

# A robust adaptive-to-model enhancement test for parametric single-index models

Cuizhen Niu<sup>1</sup> and Lixing Zhu<sup>2\*</sup>

<sup>1</sup>*School of Statistics, Renmin University of China, Beijing*

<sup>2</sup>*Department of Mathematics, Hong Kong Baptist University, Hong Kong*

**Abstract:** In the research on checking whether the underlying model is of parametric single-index structure with outliers in observations, the purpose of this paper is two-fold. First, a test that is robust against outliers is suggested. The Hampel's second-order influence function of the test statistic is proved to be bounded. Second, the test fully uses the dimension reduction structure of the hypothetical model and automatically adapts to alternative models when the null hypothesis is false. Thus, the test can greatly overcome the dimensionality problem and is still omnibus against general alternative models. The performance of the test is demonstrated by both Monte Carlo simulation studies and an application to a real dataset.

*Keywords:* Bounded influence function; Dimension reduction; Model checking; Omnibus property; Robust adaptive-to-model test.

---

\*Corresponding author: Lixing Zhu, email: lzhu@hkbu.edu.hk. The research described herewith was supported by a grant from the University Grants Council of Hong Kong and the Outstanding Innovative Talents Cultivation Funded Programs 2014 of Renmin University of China.

# 1 Introduction

Testing the validity of a specified model is an important issue in statistical inference. A long-standing focus for this model checking problem is their sensitivity to outlying observations or heavy-tailed distributions, which may have destructive effects even though with small violation in usual observations. However, in the past decades, researchers have made more effort on robust estimation whereas paid less attention to robust hypothesis testing. Therefore, it is critical to develop a robust test that can be against outlier contamination.

Outliers or contamination data are a ubiquitous problem in many disciplines, for example, clinical trials, medical research, longitudinal studies and so forth. When there exist outliers in the data, robust statistical inference procedures are necessary to improve the accuracy and reliability of results. The purpose of robust testing is two-fold, just as stated in Heritier and Ronchetti (1994): One is that under small and arbitrary departures from the null hypothesis, the level of a test should be stable, which is called the robustness of validity; The other is that the test can still make a good power performance under small and arbitrary departures from specified alternatives, that is called the robustness of efficiency. Wang and Qu (2007) suggested a robust version of Zheng's (1996) test. Their numerical studies also showed the necessity of using a robust testing procedure: the effect of outliers on Zheng's original test is dramatic and destructive so that it can not maintain the significance level.

Many efforts have been devoted to the development of robust testing procedures. For linear regression models, Schrader and Hettmansperger (1980) proposed the  $\rho_c$  test based on Huber's M estimations; Markatou and Hettmansperger (1990) introduced an aligned generalized M test for testing subhypotheses in general linear models, which is a robustification of the well known  $F$  test and can be viewed as a generalization of Sen's (1982)  $M$  test for linear models. Afterwards, Heritier and Ronchetti (1994) and Markatou and Manos (1996) presented robust versions of the Wald, score and drop-in-dispersion tests for general parametric models and nonlinear regression models, respectively. Wang and Qu (2007) developed a robust approach for testing the parametric form of a regression function versus an omnibus

alternative, which can be viewed as a robustification of a smoothing-based conditional moment test. Feng et al (2015) recommended a robust testing procedure to make comparison of two regression curves through combining a Wilcoxon-type artificial likelihood function with generalized likelihood ratio test.

There are a number of proposals available in the literature on testing consistently the correct specification of a parametric regression model. Most of existing test procedures can be classified into two categories: global smoothing tests and local smoothing tests. As mentioned in a comprehensive review paper of González-Manteiga and Crujeiras (2013), the global smoothing tests mainly involve empirical process, which can avoid subjective selection of the smoothing parameter, such as bandwidth; The local smoothing tests are based on nonparametric smoothing techniques such as Nadaraya-Waston kernel estimation (Nadaraya 1964; Waston 1964), smoothing spline estimation or other local smoothing techniques. Examples of global smoothing tests include Bierens (1990), Bierens and Ploberger (1997), Stute (1997), Stute et al (1998b), Whang (2000), Escanciano (2006a), among many others. This class of methods enjoy fast convergence rate of order  $O(n^{-1/2})$  (see Stute et al (1998a)). However, in high-dimensional scenarios, the limiting null distribution is usually intractable which requires the assistance from resampling approximation to determine critical values and is not sensitive to high-frequency models. Further, the power performance in high-dimensional cases is not very encouraging. This problem makes greatly practical limitation since it is not uncommon to have high-order or high-frequency models.

As for local smoothing tests, examples include the tests suggested by Härdle and Mammen (1993), Zheng (1996), Fan et al (2001), Horowitz and Spokoiny (2001), Koul and Ni (2004), Van Keilegom et al (2008). Since they must involve multivariate nonparametric function estimation procedures, and thus inevitably suffer from the curse of dimensionality when the number of covariates is large, even moderate. This typical problem is a big obstacle for the tests to well maintain the significance level and sense the alternative models. Because of the data sparseness in multidimensional spaces, the behavior of nonparametric smooth estimators quickly deteriorates as the dimension increases, see Stone (1980). Further, even when there are no

outliers, local smoothing tests have the typical convergence rate of order  $O(n^{-1/2}h^{-p/4})$  to their limits, which is very slow when  $p$  is large where  $p$  is the number of covariates. Besides, a suitable choice of smooth parameter is difficult but necessary for these tests. Although previous simulation and empirical studies show that the effect of bandwidth selection is not too profound for small  $p$  situation, how to make an optimal bandwidth choice is not solved thoroughly when the dimension  $p$  is relatively large.

The above analysis shows that there exists a common problem in both procedures, the global and the local, that is, the sparseness of data in high-dimensional spaces makes most of test statistics suffer curse of dimensionality, even for large sample sizes (see Escanciano (2007)). To attack this challenge, a representative method documented as projection-pursuit technique was proposed and experimented. The significant feature of this method is to employ the projection of original data onto one-dimensional subspaces: first projecting the original high-dimensional covariates to one-dimensional space to form a linear combination and a test can be obtained as an average of tests based on these selected combinations, see Huber (1985) for detail. Escanciano (2006b) proposed a consistent test for the goodness-of-fit of parametric regression models, which applied a residual marked empirical process based on projections to bypass the curse of dimensionality caused by the fact that high-dimensional space is mostly empty. Zhu and Li (1998) suggested to use projection pursuit technique to define a test that is based on an unweighed integral of expectations with respect to all one-dimensional directions. Zhu and An (1992) had already used this idea to deal with a relevant testing problem. Zhu (2003) constructed a lack-of-fit test via seeking for a good projection direction for plotting to achieve the dimension-reduction aim. Lavergne and Patilea (2008) introduced the projection-pursuit technique to local smoothing-based tests to avoid the effect of dimension. Afterwards, Lavergne and Patilea (2012) suggested a smooth integrated conditional moment (ICM) test, which is an omnibus test based on the kernel estimation that performs against a sequence of directional nonparametric alternatives as if there were only one regressor whatever the number of regressors. However, all of tests require resampling/bootstrap to determine the critical values, which is compute-intensive and time-consuming. Stute and Zhu (2002)

simply used the one-dimensional projected covariate that is based on the null model and thus the problem of dimensionality is greatly alleviated. But the disadvantage is that it is a directional rather than an omnibus test which cannot detect general alternatives. Another relevant reference is Stute, Xu and Zhu (2008) who also suggested a dimension reduction test that is based on the residual empirical process marked by a set of functions of the covariates. This test relies solely on selecting proper functions for the significance level maintainance and power enhancement.

Recently, a dimension-reduction model-adaptive test is proposed by Guo et al (2015), which is an omnibus test against global alternative models. This test introduces a model-adaptation concept in model checking for parametric regression models. The test statistic under the hypothetical model can converge to its limit at the rate of order  $O(n^{-1/2}h^{-1/4})$  and detect local alternatives distinct from the null model at this rate, which is not affected by the dimension of covariates. Their test behaves like a local smoothing test, as if the covariates were one-dimensional. Another superiority is that it owns tractable limiting null distribution and can work very well even with moderate sample sizes without the assistance of resampling approximation to determine critical values.

All of the above tests can avoid the curse of dimensionality to some extent, however, they are not robust against outliers and their efficiency is adversely affected by outlying observations. Our subsequent numerical analysis suggests that the test proposed by Guo et al (2015) is failure when there exist some outliers because a linear local average of the response variable is not robust, as elaborated in Härdle (1992). To address this problem, we intend to incorporate the idea in Guo et al (2015) into our robust model-adaptive smoothing-based conditional moment test so that it can possess the robustness property and simultaneously solve the problem of dimensionality.

The hypothetical model is the following with a dimension-reduction structure:

$$Y = g(\beta^\top X, \theta) + e, \quad (1.1)$$

where  $Y$  is the response with the covariate  $X \in \mathbb{R}^p$ , the error  $e$  is with zero mean and is independent of  $X$ .  $\beta$  and  $\theta$  are unknown parameter vectors of dimensions  $p$  and  $d$ , respectively. In addition,  $g(\cdot)$  is a known function

and the superscript  $\top$  denotes transposition. As we often have no much information in advance on model structure under alternative hypothesis, a general alternative model is considered as follows:

$$Y = m(B^\top X) + \varepsilon, \quad (1.2)$$

where  $m(\cdot)$  is an unknown smooth function and  $E(\varepsilon|X) = 0$ .  $B$  is a  $p \times q$  matrix with  $q$  orthogonal columns for an unknown number  $q$  with  $1 \leq q \leq p$ . This model treats the nonparametric regression  $E(Y|X) = m(X)$  as a special case for which the matrix  $B = I_p$  with  $q = p$ :

$$Y = m(X) + \varepsilon,$$

where  $m(\cdot)$  is an unknown smooth link function and  $E(\varepsilon|X) = 0$ .

In this paper, we construct a robust dimension reduction adaptive-to-model test (RDREAM). It sufficiently invokes the information in both the null and alternative models to get rid of curse of dimensionality and employs centered asymptotic rank transformation technique to achieve the goal of robustness. We further study the local robustness via influence function analysis, which indicates that our RDREAM has first-order influence function of zero and second-order influence function bounded in the response direction. Therefore, it is verified that our test can make more stable and robust performance when there are outliers in responses than existing local smoothing tests.

The rest of this article is organized as follows. In Section 2, the test is constructed. The approaches to estimate the matrix  $B$  and to identify its structure dimension  $q$  are also stated in this section. Section 3 presents the large sample properties under the null, global and local alternative hypothesis. Section 4 discusses the local robustness property through the Hampel influence function analysis. Numerical studies including simulation studies and a HIV real data analysis are respectively reported in Section 5 and Section 6. All of the proofs are relegated to the Appendix.

## 2 A robust dimension reduction adaptive-to-model test

As discussed before, the hypotheses of interest are:

$$\begin{aligned} H_0 : E(Y|X) &= g(\beta^\top X, \theta) \text{ for some } \beta \in R^p, \theta \in R^d; \\ H_1 : E(Y|X) &= m(B^\top X) \neq g(\beta^\top X, \theta) \text{ for any } \beta \in R^p, \theta \in R^d, \end{aligned} \quad (2.1)$$

where  $g(\cdot)$  is a known link function.  $\beta$  and  $\theta$  are respectively the parameter vectors of  $p$  and  $d$  dimensions.  $B$  is a  $p \times q$  orthonormal matrix where  $B^\top B = I_q$  and  $1 \leq q \leq p$ .

### 2.1 Test statistic construction

The key idea of the local smoothing-based conditional moment test is to apply the centered rank-transformed residuals. Denote  $e = Y - g(\beta^\top X, \theta)$  and let  $e_i^* = H(e_i) - \frac{1}{2}$ , where  $H(\cdot)$  is the distribution function of  $e$  and  $H(e_i)$ ,  $i = 1, \dots, n$  follow a uniform distribution on  $(0, 1)$ . Under the null hypothesis  $H_0$ ,

$$E(e_i^*|X_i) = E\left\{H(e_i) - \frac{1}{2}|X_i\right\} = 0, \quad (2.2)$$

In this case, the model (1.1) is with  $q = 1$ , and

$$E(e_i^*|X_i) = 0 = E(e_i^*|\beta^\top X_i) = E(e_i^*|B^\top X_i) = 0.$$

Further, the following formula

$$E\{e_i^* E(e_i^*|B^\top X_i) f(B^\top X_i)\} = E\{E^2(e_i^*|B^\top X_i) f(B^\top X_i)\} = 0 \quad (2.3)$$

holds under  $H_0$ , where  $f(\cdot)$  is the probability density function of  $B^\top X_i$ .

Under  $H_1$ , we have  $e = Y - g(\beta^\top X, \theta) = m(B^\top X) - g(\beta^\top X, \theta) + \varepsilon$ . It is easy to see that

$$\begin{aligned} & E(e_i^*|X_i) \\ = & E\left\{Q\left(\varepsilon_i + (m(B^\top X_i) - g(\beta^\top X_i, \theta)) - (m(B^\top X) - g(\beta^\top X, \theta))\right) - \frac{1}{2}|X_i\right\}, \end{aligned} \quad (2.4)$$

where  $Q(\cdot)$  is the distribution function of  $\varepsilon$ . The conditional expectation  $E(e_i^*|X_i)$  equals to zero only if for any  $x$  in a set with probability 1, there is

$$E\left\{Q\left(\varepsilon_i + (m(B^\top x) - g(\beta^\top x, \theta)) - (m(B^\top X) - g(\beta^\top X, \theta))\right) - \frac{1}{2}\right\} = 0. \quad (2.5)$$

It occurs only when  $P(m(B^\top X_j) - g(\beta^\top X_j, \theta) = C) = 1$  for some constant  $C$ , which only holds under the null hypothesis  $H_0$ . Since we can enlarge the null class of models by including some location shifts, if  $g(\beta^\top x, \theta)$  belongs to the null class of models, then so does  $g(\beta^\top x, \theta) + C$ ; in other words, it is reasonable to assume that the null class of models is sufficiently general to contain all location shifts in the  $y$  direction. Therefore, under  $H_1$ , the formula (2.5) does not hold. Further, based on (2.4), the conditional expectation  $E(e_i^*|X_i)$  is not zero and

$$E(e_i^*|X_i) \neq 0 \Leftrightarrow E(e_i^*|B^\top X_i) \neq 0.$$

Thus, we have

$$E\{e_i^* E(e_i^*|B^\top X_i) f(B^\top X_i)\} = E\{E^2(e_i^*|B^\top X_i) f(B^\top X_i)\} > 0. \quad (2.6)$$

Based on the different performance of  $E\{e_i^* E(e_i^*|B^\top X_i) f(B^\top X_i)\}$  under  $H_0$  and  $H_1$  in (2.3) and (2.6), respectively, the empirical version of it can be applied to construct a test statistic. The null hypothesis  $H_0$  is rejected for large values of the test statistic.

Given a random sample  $\{(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)\}$ , define the asymptotic rank transform of  $\hat{e}_i = y_i - g(\hat{\beta}^\top x_i, \hat{\theta})$  as  $n^{-1} \sum_{l=1}^n I(\hat{e}_l \leq \hat{e}_i)$  where  $\sum_{l=1}^n I(\hat{e}_l \leq \hat{e}_i)$  is the rank of  $\hat{e}_i$  among all of the  $n$  residuals. Here,  $\hat{\beta}$  and  $\hat{\theta}$  come from robust estimates of  $\beta$  and  $\theta$ . Further, the corresponding centered asymptotic rank transform of residuals is as follows

$$\hat{e}_i^* = \frac{1}{n} \sum_{l=1}^n I(\hat{e}_l \leq \hat{e}_i) - \frac{n+1}{2n}, \quad l = 1, \dots, n, \quad (2.7)$$

where  $I(\cdot)$  is the indicator function.

Once an estimator of  $\hat{B}(\hat{q})$  is available, a kernel estimator of the regression function  $E(e_i^*|B^\top X_i)$  can be estimated as follows:

$$\hat{E}(e_i^*|\hat{B}(\hat{q})^\top x_i) = \frac{\frac{1}{n-1} \sum_{j \neq i} \hat{e}_j^* \mathcal{K}_h\{\hat{B}(\hat{q})^\top (x_i - x_j)\}}{\hat{f}(\hat{B}(\hat{q})^\top x_i)},$$



where  $\hat{e}_j^*$  has been defined in (2.7).  $\hat{B}(\hat{q})$  is an sufficient dimension reduction (SDR) estimate of the matrix  $B$  with an estimated structural dimension  $\hat{q}$  of  $q$  and the estimates will be specified later. Besides,  $\mathcal{K}_h(\cdot) = \mathcal{K}(\cdot/h)/h^{\hat{q}}$ , where  $\mathcal{K}(\cdot)$  is a  $\hat{q}$ -dimensional kernel function and  $h$  is the bandwidth, and  $\hat{f}(\hat{B}(\hat{q})^\top x_i)$  is a kernel estimator of the density function of  $f(B^\top x_i)$ ,

$$\hat{f}(\hat{B}(\hat{q})^\top x_i) = \frac{1}{n-1} \sum_{j \neq i}^n \mathcal{K}_h\{\hat{B}(\hat{q})^\top (x_i - x_j)\}.$$

Further, a robust dimension reduction adaptive-to-model test (RDREAM) can be constructed as follows:

$$V_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K}_h\{\hat{B}(\hat{q})^\top (x_i - x_j)\} \hat{e}_i^* \hat{e}_j^*. \quad (2.8)$$

**Remark 1.** From the above construction of  $V_n$  in (2.8), it seems that except for the estimates of the matrix  $B$  and structural dimension  $q$ , the test statistic makes no difference with the test proposed by Wang and Qu (2007) as follows:

$$\tilde{V}_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \tilde{\mathcal{K}}_h(x_i - x_j) \hat{e}_i^* \hat{e}_j^*, \quad (2.9)$$

where  $\tilde{\mathcal{K}}_h(\cdot) = \tilde{\mathcal{K}}(\cdot/h)/h^p$  with  $\tilde{\mathcal{K}}(\cdot)$  being a  $p$ -dimensional kernel function. Comparing the test statistic in (2.8) with that in (2.9), we note that  $\hat{q}$ -dimensional kernel function ( $\hat{q} \leq p$ ) is required in  $V_n$ . The result in Section 3 shows that under the null hypothesis  $H_0$ ,  $\hat{q} \rightarrow 1$ , which can avoid the curse of dimensionality greatly. Another superiority of the new test is the model-adaptive property, that is, through estimating the matrix  $B$ , the test can automatically adapt the hypothetical and alternative model such that it can have better performance in the significance level maintainance and power enhancement. To be specific, under  $H_0$ ,  $\hat{B}(\hat{q}) \rightarrow c\beta$  for a constant  $c$ , and under  $H_1$   $\hat{q} \rightarrow q \geq 1$ ,  $\hat{B}(\hat{q}) \rightarrow BC$  for a  $q \times q$  orthogonal matrix, adaptive to the alternative model (1.2).

**Remark 2.** *Another related test is the dimension reduction model-adaptive test  $T_n^{GWZ}$  proposed by Guo et al. (2015):*

$$T_n^{GWZ} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K}_h\{\hat{B}(\hat{q})^\top (x_i - x_j)\} \hat{e}_i \hat{e}_j. \quad (2.10)$$

*In the Section 4, we will show through Hampel influence function analysis that RDREAM has more stable and robust behavior than  $T_n^{GWZ}$  when the response is contaminated.*

## 2.2 Identification and estimation of $B$

As the estimates of the matrix  $B$  and structural dimension  $q$  are crucial for our RDREAM, we first specify the estimate for the matrix  $B$  under given  $q$  and then study how to select  $q$  consistently. Note that the model (1.2) is a multi-index model with unknown  $q$  indexes, thus outer product of gradients (OPG) introduced by Xia et al (2002) can be considered to estimate  $B$ . Another method to estimate  $B$  is inspired by sufficient dimension reduction technique. In fact,  $B$  is not identifiable since for any  $q \times q$  orthogonal matrix  $C$ ,  $m(B^\top X)$  can always be rewritten as  $\tilde{m}(C^\top B^\top X)$ . Therefore, what we can identify is the space spanned by  $B$  via sufficient dimension reduction technique, or in other words, we can identify  $q$  base vectors of the space spanned by  $B$ . There exist several proposals in the literature, such as sliced inverse regression (SIR) proposed by Li (1991), sliced average variance estimation (SAVE) considered by Cook and Weisberg (1991), minimum average variance estimation (MAVE) advised by Xia et al (2002) and discretization-expectation estimation (DEE) suggested by Zhu et al (2010). In view of easy-operation and good-performance of DEE, we consider to employ it to estimate  $B$ . Further, as the method called outer product of the gradients in Xia et al (2002) has less restriction on the covariates  $X$ , we then also use it to estimate  $B$  for a comparison with SIR-based DEE. In the following, we give simple review of OPG and SIR-based DEE.

### 2.2.1 Outer product gradients

The outer product of the gradients can be written as

$$E\{\nabla m(B^\top X)\nabla m(B^\top X)^\top\} = BE\{m'(B^\top X)m'(B^\top X)^\top\}B^\top,$$

where  $m(B^\top x) = E(Y|X = x)$ ,  $\nabla m(B^\top X) = \frac{\partial}{\partial X}m(B^\top X)$  and  $m'(B^\top X) = \frac{\partial}{\partial B^\top X}m(B^\top X)$ . Note that  $E\{\nabla m(B^\top X)\nabla m(B^\top X)^\top\}$  has  $q$  nonzero eigenvalues, the matrix  $B$  is in the space spanned by the  $q$  eigenvectors corresponding to the largest  $q$  eigenvalues of  $E\{\nabla m(B^\top X)\nabla m(B^\top X)^\top\}$ . Thus, we need to estimate the expectation  $E\{\nabla m(B^\top X)\nabla m(B^\top X)^\top\}$  and then obtain the estimator of  $B$ .

By local linear fitting

$$m(B^\top x_i) = m(B^\top x_j) + m'(B^\top x_j)^\top B^\top (x_i - x_j) = a_j + b_j^\top x_{ij},$$

where  $a_j = m(B^\top x_j)$ ,  $b_j = B \times m'(B^\top x_j)$  and  $x_{ij} = x_i - x_j$ , we can obtain  $(\hat{a}_j, \hat{b}_j)$  by minimizing the following objective function

$$\min_{a_j, b_j} \sum_{i=1}^n \mathcal{K}_h(B^\top x_{ij}) \{y_i - a_j - b_j^\top x_{ij}\}^2, \quad (2.11)$$

where  $\mathcal{K}_h(\cdot) = \mathcal{K}(\cdot/h)/h$  with  $\mathcal{K}(\cdot)$  being a  $q$ -dimensional kernel function and  $h$  being a bandwidth. The corresponding estimating equation from (2.11) for  $(a_j, b_j)$  can be given as

$$\sum_{i=1}^n \mathcal{K}_h(B^\top x_{ij})(1, x_{ij}^\top)^\top \{y_i - \hat{a}_j - \hat{b}_j^\top x_{ij}\} = 0.$$

Further,  $E\{\nabla m(B^\top X)\nabla m(B^\top X)^\top\}$  can be estimated as

$$\hat{\Sigma} = \frac{1}{n} \sum_{j=1}^n \hat{b}_j \hat{b}_j^\top. \quad (2.12)$$

Therefore, the  $q$  eigenvectors corresponding to the largest  $q$  eigenvalues of  $\hat{\Sigma}$  can be regarded as the estimator of the matrix  $B$ .

### 2.2.2 Discretization-expectation estimation

We first give some notations. The central subspace, denoted by  $\mathcal{S}_{Y|X}$ , is defined as the intersection of all subspaces spanned by the columns of a matrix  $A$ ,  $\mathcal{S}_A$ , of minimal dimension such that  $Y \perp\!\!\!\perp X|A^\top X$ , where  $\perp\!\!\!\perp$  stands for statistical independence. Similarly, let  $\mathcal{S}_{E(Y|X)}$  be the intersection of all subspaces  $\mathcal{S}_A$  spanned by the matrix  $A$  such that  $Y \perp\!\!\!\perp E(Y|X)|A^\top X$ . In sufficient dimension reduction,  $\mathcal{S}_{E(Y|X)}$  is called the central mean subspace and its dimension, denoted by  $d_{E(Y|X)}$ , is called the structural dimension. As to the model (1.2), we have  $\mathcal{S}_{E(Y|X)} = \text{span}(B)$  and  $d_{E(Y|X)} = q$ . Thus, we aim to identify the  $q$  base vectors of  $\mathcal{S}_{E(Y|X)}$ .

Compared with SIR, DEE can avoid the choice of the number of slices, as Li and Zhu (2007) pointed out, which may affect the efficiency and even lead to inconsistent estimates, and has no optimal solution. Define the new response variable  $Z(t) = I(Y \leq t)$  for any  $t$ , where the indicator function  $I(Y \leq t)$  takes the value 1 if  $Y \leq t$  and 0, otherwise. When SIR is applied, the original related matrix  $\mathcal{M}(t)$  based on SIR is a  $p \times p$  positive semi-definite matrix such that  $\text{span}\{\mathcal{M}(t)\} = \mathcal{S}_{Z(t)|X}$ . Here,  $\mathcal{S}_{Z(t)|X}$  is the central subspace of  $Z(t)|X$ . Given  $\mathcal{M} = E\{\mathcal{M}(T)\}$ , according to Theorem 1 in Zhu et al (2010),  $\text{span}\{\mathcal{M}\} = \mathcal{S}_{Y|X}$ . To ensure  $\mathcal{S}_{Y|X} = \mathcal{S}_{E(Y|X)}$  for  $\varepsilon$  in the model (1.2), based on Guo et al (2015), a condition that  $\varepsilon = m_1(B^\top X)\tilde{\varepsilon}$  and  $\tilde{\varepsilon} \perp\!\!\!\perp X$  is needed.

Based on the above analysis, estimating  $\mathcal{S}_{E(Y|X)}$  amounts to estimating  $\mathcal{M}$ . Given the sample  $\{(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)\}$ , we define the dichotomized responses as  $z_i(y_j) = I(y_i \leq y_j)$ ,  $i, j = 1, \dots, n$ . Thus, for each fixed  $y_j$ , we can obtain a new sample  $\{(z_1(y_j), x_1), (z_2(y_j), x_2), \dots, (z_n(y_j), x_n)\}$  and the estimate  $\mathcal{M}_n(y_j)$  of  $\mathcal{M}(y_j)$  can be gained with SIR. Thus,  $\mathcal{M}$  can be estimated as

$$\mathcal{M}_{n,n} = n^{-1} \sum_{j=1}^n \mathcal{M}_n(y_j), \quad (2.13)$$

which has been proved to be root- $n$  consistent to  $\mathcal{M}$  by Zhu et al (2010). The  $q$  eigenvectors of  $\mathcal{M}_{n,n}$  corresponding to its  $q$  largest eigenvalues are applied to estimate  $B$ . For this method, a mild linearity condition is assumed:  $E(X|B^\top X = u)$  is linear in  $u$  (Li, 1991).

### 2.3 Estimation of structural dimension $q$

In order to get RDREAM in (2.8), the estimate of structural dimension  $q$  is necessary for the above two methods of identifying  $B$ . Here, a Ridge-type Ratio Estimate (RRE) method, which is inspired by Xia et al (2015), is suggested to determine  $q$  for OPG and DEE. It is based on the ratios of the eigenvalues with an artificially added ridge value  $c$ . Denote  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$  to be the eigenvalues of the estimating matrix  $\hat{\Sigma}$  or  $\mathcal{M}_{n,n}$  in (2.12) and (2.13), respectively.  $\hat{q}$  can be obtained as

$$\hat{q} = \arg \min_{k=1, \dots, p-1} \frac{\lambda_{k+1} + c}{\lambda_k + c},$$

where the constant  $c = 1/\sqrt{nh}$  is recommended. The consistencies of  $\hat{q}$  under the null hypothesis (1.1) and global alternative hypothesis (1.2) are shown in the following lemma.

**Lemma 1.** *Assume that the OPG-based matrix  $\hat{\Sigma}$  or the DEE-based matrix  $M_{nn}$  is root- $n$  consistent to  $\Sigma$  or  $M$ . Then the corresponding estimate  $\hat{q} = q$  as  $n \rightarrow \infty$  with a probability going to one. Therefore, for a  $q \times q$  orthogonal matrix  $C$ ,  $\hat{B}(\hat{q})$  is a root- $n$  consistent estimate of  $BC^\top$ .*

## 3 Asymptotic properties

In this section, the large-sample properties of the RDREAM test statistic  $V_n$  in (2.8) are investigated via its asymptotic distributions under the null hypothesis, global alternative hypothesis and local alternative hypothesis.

### 3.1 Limiting null distribution

The asymptotic normality discussed in the following also requires that the regression parameter is root- $n$  consistently estimated under  $H_0$  and the residuals must come from a robust fit. Let  $Z = B^\top X$  and

$$Var = \frac{1}{72} \int \mathcal{K}^2(u) du \int p^2(z) dz,$$

where  $p(z)$  is the probability density function of  $Z$ . Moreover,  $Var$  can be consistently estimated by:

$$\widehat{Var} = \frac{1}{72n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^q} \mathcal{K}^2 \left\{ \frac{\hat{B}(\hat{q})^\top (x_i - x_j)}{h} \right\}.$$

We first state the asymptotic property of the RDREAM test statistic in (2.8) under the null hypothesis  $H_0$  as follows:

**Theorem 1.** *Suppose that conditions (C1)-(C8) in the Appendix hold. Under  $H_0$ , we have*

$$nh^{1/2}V_n \Rightarrow N(0, Var).$$

Plugging in a consistent estimator of  $Var$ , a standardized version of  $V_n$  can be defined as

$$S_n = \frac{n-1}{n} \frac{nh^{1/2}V_n}{\sqrt{\widehat{Var}}}. \quad (3.1)$$

The following corollary can be easily obtained.

**Corollary 1.** *Under  $H_0$  and Conditions (C1)-(C8) in the Appendix, we have*

$$S_n^2 \Rightarrow \chi_1^2,$$

where  $\chi_1^2$  is the chi-square distribution with one degree of freedom.

Theorem 1 and Corollary 1 characterize the asymptotic properties of the test statistic  $V_n$ . Based on Corollary 1,  $p$ -values of RDREAM can be easily determined by the quantiles of the chi-square distribution with one degree of freedom. The null hypothesis  $H_0$  is rejected when  $S_n \geq \chi_{1-\alpha}^2(1)$  where  $\chi_{1-\alpha}^2(1)$  is the  $1 - \alpha$  upper quantile of the chi-square distribution.

### 3.2 Power study

We are now in the position to examine the power performance of our RDREAM under alternative hypothesis. More specifically, the following sequence of alternative models is under consideration:

$$H_{1n} : \quad Y = g(\tilde{\beta}^\top X, \tilde{\theta}) + C_n m(B^\top X) + \varepsilon, \quad (3.2)$$

where  $E(\varepsilon|X) = 0$ ,  $E[m^2(B^\top X)] < \infty$  and  $\{C_n\}$  is a constant sequence. When  $C_n = C$  for a nonzero constant  $C$ , the model is a global alternative model, while when  $C_n$  goes to zero, it is a sequence of local alternative models. In this sequence of models,  $\beta$  is one of the column in  $B$ . Denote  $\tilde{\alpha} = (\beta, \theta)$  to be

$$\tilde{\alpha} = \arg \min_{\alpha} E\{g(\beta^\top X, \theta) - m(X)\},$$

where  $m(X) = E(Y|X)$ . When the null hypothesis  $H_0$  holds,  $\tilde{\alpha}$  is the true parameter. For a robust estimate  $\hat{\alpha}$ , we have  $\hat{\alpha} - \tilde{\alpha} = O_p(1/\sqrt{n})$ . We first discuss the consistency of  $\hat{q}$  under the local alternative hypothesis (3.2). When  $n \rightarrow \infty$ , the local alternative models converge to the null model,  $\hat{q}_s$  under the local alternative models are expected to converge to  $\hat{q}$  under the null model, which finally converge to the structural dimension  $q = 1$  under the null model.

**Lemma 2.** *Assume conditions (C1)-(C8) in the Appendix hold and under the local alternative hypothesis (3.2) with  $C_n = n^{-1/2}h^{-1/4}$ , we have  $\hat{q} = 1$  as  $n \rightarrow \infty$  with a probability going to one, where  $\hat{q}$  is either the OPG-based or the DEE-based estimate.*

The asymptotic properties under global and local alternative hypotheses are concluded in the following Theorem.

**Theorem 2.** *Under Conditions (C1)-(C8) in the Appendix, we have:*

(i) *Under the global alternative of (1.2) or equivalently the above model with  $C_n = C$ ,*

$$S_n/(nh^{1/2}) \Rightarrow \text{Constant} > 0;$$

(ii) *Under the local alternative hypothesis (3.2) with  $C_n = n^{-1/2}h^{-1/4}$ , we have*

$$nh^{1/2}V_n \Rightarrow N(\mu, \text{Var}) \quad \text{and} \quad S_n^2 \Rightarrow \chi_1^2(\mu^2/\text{Var}),$$

where

$$\mu = E[h^2(\varepsilon)m^2(B^\top X)p(X)],$$

$h(\cdot)$  denotes the probability density function of  $\varepsilon$  and  $\chi_1^2(\mu^2/\text{Var})$  is a non-central chi-squared random variable with one degree of freedom and the non-centrality parameter  $\mu^2/\text{Var}$ .

The above Theorem indicates that under the global alternative hypothesis, the RDREAM test is consistent with the asymptotic power 1 and can detect the local alternatives distinct from the null hypothesis at a non-parametric rate of order  $n^{-1/2}h^{-1/4}$ , which is the optimal rate with one dimensional predictor for the test  $\tilde{V}_n$  in (2.9).

**Remark 3.** From the above theorem, we can observe that under the null hypothesis, the use (automatically) of lower order kernel function in  $V_n$  in (2.8) makes a very significant improvement than  $\tilde{V}_n$  in (2.9). Based on Theorem 1,  $V_n$  owns a much faster convergence rate of order  $nh^{1/2}$  and  $nh^{1/2}V_n$  is asymptotically normal under the null whereas the rate of order  $nh^{p/2}$  is for  $\tilde{V}_n$ . Further, according to Theorem 2, the conclusion can be made that  $V_n$  is much more sensitive than Wang and Qu's test  $\tilde{V}_n$  in the sense that  $V_n$  can detect the local alternatives distinct from the null at the rate of  $n^{-1/2}h^{-1/4}$  whereas  $\tilde{V}_n$  is only workable at the rate of order  $n^{-1/2}h^{-p/4}$ . Therefore, the power performance of the proposed test can be much enhanced.

**Remark 4.** Our original simulation results based on  $S_n$  in (3.1) suggest the conservative sizes of tests. Thus, the following size-adjustment is needed for the test statistics with both the OPG-based estimate and the DEE-based estimate:

$$\tilde{S}_n = (1 + 4n^{-4/5})S_n. \quad (3.3)$$

The size-adjustment constant is chosen through intensive simulation with various different values and this one is recommended. With such a size-adjustment, our new test  $\tilde{S}_n$  can better control type I errors. It is worth noting that the size-adjustment is asymptotically negligible when  $n \rightarrow \infty$  since  $\tilde{S}_n \rightarrow S_n$  when  $n \rightarrow \infty$ .



## 4 Robustness property

In this section, we investigate the local stability of RDREAM under infinitesimal local contamination through Hampel influence function, which was introduced by Hampel (1974). The Hampel influence analysis reveals that RDREAM has the desired local robustness property. In the following, we first give von Mises functional expansion of RDREAM and further derive the Hampel influence function. For this investigation, Wang and Qu (2007) is a good reference. The key difference is between the used covariate  $X$  and  $\hat{B}(\hat{q})^\top X$  in the respective test statistics, and  $\hat{B}(\hat{q})^\top X$  is automatically 1- and  $q$ -dimensional under the null and alternative hypothesis. Thus, we only give some brief descriptions about the results.

Following Wang and Qu (2007), the von Mises analysis (see Fernholz (1983)) can provide the basis of the Hampel influence function calculation. We now discuss the von Mises functional expansion.

Let  $Z = B^\top X$ ,  $\hat{Z} = \hat{B}(\hat{q})^\top X$  and denote  $\hat{H}_n$  as the empirical distribution function of  $Y_1 - g(\hat{\beta}^\top X_1, \hat{\theta}), \dots, Y_n - g(\hat{\beta}^\top X_n, \hat{\theta})$ . When  $H_0$  holds,  $B = c\beta$ , and  $\hat{B}(\hat{q})$  is an estimator of  $\beta$  up to a scalar constant  $c$ . The RDREAM statistic  $V_n$  in (2.8) can be asymptotically expressed as

$$T(\hat{F}_h, F_n) =: \int \int \left[ \hat{H}_n(y - g((\beta^\top x)(F_n), \theta(F_n))) - \frac{1}{2} \right] \times \left( \int \left[ \hat{H}_n(y_1 - g((\beta^\top x)(F_n), \theta(F_n))) - \frac{1}{2} \right] \times \hat{f}_h(\hat{z}, y_1) dy_1 \right) dF_n(\hat{z}, y),$$

where  $F_n(\cdot)$  is the empirical distribution function of  $(\hat{Z}_i, Y_i)$ 's,  $((\beta^\top x)(F_n), \theta(F_n))^\top$  is an estimator of  $(\beta^\top x, \theta)^\top$ , which can be rewritten as a functional of the empirical distribution  $F_n$ , and  $\hat{f}_h(z_i, y_i)$  is a smoothing kernel estimation of the joint density function of  $(\hat{Z}, Y)$ , which has the following form:

$$\hat{f}_h(\hat{z}_i, y_i) = \frac{1}{n-1} \sum_{i \neq j}^n \mathcal{K}_{1,h}(y_i - y_j) \mathcal{K}_{2,h}(\hat{z}_i - \hat{z}_j),$$

where  $\mathcal{K}_1(\cdot), \mathcal{K}_2(\cdot)$  are two kernel functions and satisfy  $\mathcal{K}_{1,h}(\cdot) = \mathcal{K}_1(\cdot/h)/h$  and  $\mathcal{K}_{2,h}(\cdot) = \mathcal{K}_2(\cdot/h)/h^{\hat{q}}$ , respectively. An appropriate functional for RDREAM is bivariate with the form

$$T(F, F)$$

$$\begin{aligned}
= & \int \int \left[ H(y - g((\beta^\top x)(F), \theta(F))) - \frac{1}{2} \right] \times \left( \int \left[ H(y_1 - g((\beta^\top x)(F), \theta(F))) - \frac{1}{2} \right] \right. \\
& \left. \times f(z, y_1) dy_1 \right) dF(z, y),
\end{aligned} \tag{4.1}$$

here,  $F(\cdot)$  is the distribution function of  $(Z, Y)$  and  $H(\cdot)$  is the distribution function of  $Y - g(\beta^\top X, \theta)$ :  $H(v) = \int \int_{y \leq g((\beta^\top x)(F), \theta(F)) + v} dF(z, y)$ .

Now we are in the position to analyse Hampel influence function (Hampel 1974). When the second-order influence function in the response direction, the regression functions is robustly estimated. Thus, we calculate the first-order and second-order influence function of  $T(F, F)$  in (4.1) at the point  $(z_0, y_0)$ . When  $H_0$  holds, the Hampel's first-order influence function of  $T(F, F)$  in (4.1) at the point  $(z_0, y_0)$  is defined as

$$IF^{(1)}(z_0, y_0; T) = \lim_{t \rightarrow 0} \frac{T(F_t, F_t) - T(F, F)}{t},$$

where  $T(F, F) = 0$  and  $F_t = (1 - t)F + t\Delta_{z_0, y_0}$ , here,  $\Delta_{z_0, y_0}$  is the point mass function at the point  $(z_0, y_0)$ . Denote  $L(t) = T(F_t, F_t) = T(F + tU, F + tU)$ , where  $U = \Delta_{z_0, y_0} - F$ . Thus, we have  $L(0) = 0$ . From the proof in the Appendix, it is not difficult to obtain that  $\frac{dL(t)}{dt}|_{t=0} = 0$ . Therefore, RDREAM has a degenerate first-order influence function.

To obtain the second-order influence function of RDREAM, we first compute

$$\begin{aligned}
& \frac{1}{2} \frac{d^2}{dt^2} L(t)|_{t=0} \\
= & \int \left( \int \dot{H}(y - g((\beta^\top x)(F), \theta(F))) f(z, y) dy \right)^2 dz \\
& + \int \int \dot{H}(y - g((\beta^\top x)(F), \theta(F))) \times \int \left[ H(y_1 - g((\beta^\top x)(F), \theta(F))) - \frac{1}{2} \right] \\
& \times u(z, y_1) dy_1 dF(z, y) + \int \int \left[ H(y - g(\beta^\top x(F), \theta(F))) - \frac{1}{2} \right] \\
& \times \int \dot{H}(y_1 - g((\beta^\top x)(F), \theta(F))) \times f(z, y_1) dy_1 dU(z, y) \\
& + \int \int \left[ H(y - g((\beta^\top x)(F), \theta(F))) - \frac{1}{2} \right] \times \left( \int \left[ H(y_1 - g((\beta^\top x)(F), \theta(F))) \right. \right. \\
& \left. \left. - \frac{1}{2} \right] u(z, y_1) dy_1 \right) dU(z, y),
\end{aligned} \tag{4.2}$$

where  $\dot{H}(\cdot) =: \frac{dH_t(\cdot)}{dt}|_{t=0}$  and  $H_t(\cdot)$  represents  $H(\cdot)$  under contamination, that is,  $H_t(v) = \int \int_{y \leq g((\beta^\top x)(F_t), \theta(F_t)) + v} dF_t(z, y)$ . Besides,  $u(z, y)$  is the probability density function of  $U(z, y)$ . Taking  $U = \Delta_{z_0, y_0} - F$  into the formula (4.2), it is shown that the four terms in (4.2) converge at the same rate. Further, based on Hampel's definition, we can obtain the second-order influence function of RDREAM at the point  $(z_0, y_0)$  as follows:

$$\begin{aligned} IF^{(2)}(z_0, y_0) &= \int \left( \int \dot{H}_{\Delta_{(z_0, y_0)}}(y - g((\beta^\top x)(F), \theta(F))) f(z, y) dy \right)^2 dz \\ &\quad + \left[ H(y_0 - g((\beta^\top x_0)(F), \theta(F))) - \frac{1}{2} \right] \times \\ &\quad \int \int \dot{H}_{\Delta_{(z_0, y_0)}}(y - g((\beta^\top x)(F), \theta(F))) dF(z, y) \\ &\quad + \left[ H(y_0 - g((\beta^\top x)_0(F), \theta(F))) - \frac{1}{2} \right] \times \int \dot{H}_{\Delta_{(z_0, y_0)}}(y - g(\beta^\top x_0(F), \theta(F))) \\ &\quad \times f(z_0, y) dy + \left[ H(y_0 - g((\beta^\top x)_0(F), \theta(F))) - \frac{1}{2} \right]^2, \end{aligned}$$

where  $z_0 = \beta^\top x_0$  and  $\dot{H}_{\Delta_{(z_0, y_0)}}(y - g((\beta^\top x)(F), \theta(F)))$  denotes  $\frac{d}{dt} H(y - g((\beta^\top x)(F_t), \theta(F_t)))|_{t=0, U=\Delta_{(z_0, y_0)}-F}$ . The detail expression of  $\dot{H}_{\Delta_{(z_0, y_0)}}(y - g((\beta^\top x)(F), \theta(F)))$  can be written as

$$\begin{aligned} &\dot{H}_{\Delta_{(z_0, y_0)}}(y - g((\beta^\top x)(F), \theta(F))) \\ &= \int h(y - g((\beta^\top x)(F), \theta(F))) \times \text{grad}_\alpha \{g((\beta^\top x)(F), \theta(F))\}^\top \times \frac{d\alpha}{dt} dF_X(x) \\ &\quad + I(y_0 \leq y + g((\beta^\top x)_0(F), \theta(F)) - g((\beta^\top x)(F), \theta(F))) \\ &\quad - H(y - g((\beta^\top x)(F), \theta(F))), \end{aligned}$$

where  $\alpha = (\beta, \theta)^\top$  and  $\text{grad}_\alpha \{g((\beta^\top x)(F), \theta(F))\}^\top$  represents the gradient of  $g((\beta^\top x)(F), \theta(F))$  with respect to  $\alpha$ . When the parameter  $\alpha$  comes from a robust fit, we have that  $d\alpha/dt$  is bounded. Together with the conditions (C1) and (C5) in the Appendix, it can be shown that  $\dot{H}_{\Delta_{(z_0, y_0)}}(y - g((\beta^\top x)(F), \theta(F)))$  is also bounded. Further, the second-order influence function  $IF^{(2)}(z_0, y_0)$  of RDREAM is bounded in the response direction.

For the purpose of comparison, we can similarly derive the first-order and second-order influence function for the test  $T_n^{GWZ}$  in (2.10). the first-order influence function is also zero and the second-order influence function

can be derived as

$$\begin{aligned}
IF_{GWZ}^{(2)}(z_0, y_0) &= \int \left( \frac{d}{dt} g((\beta^\top x)(F), \theta(F)) f(x) \right)^2 dz + [y_0 - g((\beta^\top x)_0(F), \theta(F))]^2 \\
&\quad - [y_0 - g((\beta^\top x)_0(F), \theta(F))] \int \frac{d}{dt} g((\beta^\top x)(F), \theta(F)) dz \\
&\quad - [y_0 - g((\beta^\top x)_0(F), \theta(F))] \frac{d}{dt} g((\beta^\top x)_0(F), \theta(F)) f(z_0),
\end{aligned}$$

where  $\frac{d}{dt} g((\beta^\top x)(F), \theta(F)) = \frac{d}{dt} g((\beta^\top x)(F_t), \theta(F_t))|_{t=0, U=\Delta_{(z_0, y_0)} - F}$ .

The second-order influence function  $IF_{GWZ}^{(2)}(z_0, y_0)$  in the  $y$ -direction is not bounded.

The above influence function analysis indicates RDREAM possesses more stable and robust performance than the test  $T_n^{GWZ}$  when the response is under contamination.

## 5 Simulation studies

In this section, three simulation studies are conducted to examine the theory and the finite-sample performance of the proposed RDREAM. Throughout this section, denote the adjusted RDREAM statistic in the formula (3.3) based on the OPG and SIR-based DEE estimate as  $\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$ , respectively. The purpose of the simulation studies is three-fold: to examine the robustness of the new method; to check the usefulness to overcome the curse of dimensionality; to demonstrate its usefulness in the cases without outliers. To this end, the objective of the first study is to check and compare the performance of RDREAM:  $\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$ . The effects of different distributions of the error and nonlinearity under the null hypothesis on the performance of the new tests is also considered in this study. The second study is used to examine the impact from dimensionality on both RDREAM and the robust test  $T_n^{WQ}$  introduced by Wang and Qu (2007). Since the test  $T_n^{GWZ}$  proposed by Guo et al (2015) is also to solve the dimensionality problem in model checking, the third study aims to show the robustness properties of the proposed test via comparing  $\tilde{S}_n^{OPG}$ ,  $\tilde{S}_n^{DEE}$  with  $T_n^{GWZ}$  and to examine its performance in the case without outliers.

*Study 1:* Consider the following models

$$\begin{aligned}
H_{11} : Y &= \beta^\top X + a \times \exp(-1.5\beta^\top X) + \varepsilon, \\
H_{12} : Y &= \beta^\top X + 1.5a(\beta^\top X)^3 + \varepsilon, \\
H_{13} : Y &= \beta^\top X + 6a \times \cos(0.8\pi\beta^\top X) + \varepsilon, \\
H_{14} : Y &= 2.5 \exp(0.5\beta^\top X) + 1.5a(\beta^\top X)^3 + \varepsilon,
\end{aligned}$$

where  $p = 8$ ,  $\beta = (1, \dots, 1)^\top / \sqrt{p}$ . The covariate  $X = (X_1, \dots, X_p)^\top$  are i.i.d and generated from a multivariate normal distribution  $N(0, I_p)$  where  $I_p$  is a  $p \times p$  identity matrix. Consider two kinds of errors: one is  $\varepsilon \sim N(0, 1)$  and the other is that  $\varepsilon_i$ 's are i.i.d from a log-normal distribution  $\ln N(0, 0.25)$  that is standardized to have mean 0 and variance 1. For all of models, 10% of the responses are randomly added by an outlying value 5. We set  $a = 0, 0.2, \dots, 1.0$  where  $a = 0$  corresponds to the null hypothesis and  $a \neq 0$  corresponds to the alternative hypothesis. For  $H_{11}, H_{12}$  and  $H_{13}$ , the null models are a linear model and the alternative models are all single-index models. The null model of  $H_{14}$  is a nonlinear model. Under the alternatives, the third model  $H_{13}$  is high-frequent and the other three models are not. We intend to examine whether the new tests can be powerful for both the two types of models.

To compute the robust test statistics  $\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$ , the residuals are obtained from a robust regression through  $M$ -estimate. As to the non-parametric regression estimation, throughout these simulations, unless otherwise specified, the kernel function is taken to be  $\mathcal{K}(u) = 15/16(1 - u^2)^2$  if  $|u| \leq 1$  and 0, otherwise. Our experience in the simulations suggests that RDREAM is not sensitive to the choice of kernel function. The bandwidth is recommended as  $h_{OPG} = 1.8n^{-1/(\hat{q}+4)}$  for the test statistic  $\tilde{S}_n^{OPG}$  and  $h_{DEE} = 0.5n^{-1/(\hat{q}+4)}$  for  $\tilde{S}_n^{DEE}$  through intensive numerical computation. The significance level is set to be  $\alpha = 0.05$  and the sample size  $n = 60, 100, 200$  are considered. Every simulation result is the average of 2000 replications.

Table 1 displayed the empirical sizes and powers of the new tests against the alternatives  $H_{11}, H_{12}$  with different values of  $a$ . From this table, we can see that for every case we conduct, the adjusted test statistics  $\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$  can maintain the significance level very well even with the sample

size  $n = 60$ . It is reasonable that the larger the sample size is, the closer the empirical size is to the significance level. As to the power performance, for any a specific test  $\tilde{S}_n^{OPG}$  or  $\tilde{S}_n^{DEE}$ , the powers are higher with larger sample sizes.  $\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$ , from the results of  $H_{11}$ , it can be seen that in most cases,  $\tilde{S}_n^{OPG}$  outperforms  $\tilde{S}_n^{DEE}$ ; for  $H_{12}$ ,  $\tilde{S}_n^{DEE}$  exhibits slightly higher powers than  $\tilde{S}_n^{OPG}$ , however, the difference between them can be negligible. Therefore, the test  $\tilde{S}_n^{OPG}$  yields to no worse power performance than  $\tilde{S}_n^{DEE}$ . At last, for every specific alternative model, the impact of error distribution on empirical sizes and powers shows no significant difference, which, to some extent, illustrates that the proposed tests are robust to light and heavy tail error distributions.

Table 1 about here

To get a sense of the performance under high-frequent alternative model and the effect of nonlinear null model on the proposed tests, we conduct a more in-depth analysis for the alternatives  $H_{13}$  and  $H_{14}$ . The simulated power curves of different values of  $a$  and sample sizes  $n = 100, 200$  are displayed in Figure 1. From this figure, we can see that for the cosine alternative  $H_{13}$ , the simulated power curve for  $\tilde{S}_n^{OPG}$  has a sigmoidal shape, whereas the curve for  $\tilde{S}_n^{DEE}$  shows rapid growth at the beginning and slightly decrease for a relative larger  $a$ . The different power performance for  $\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$  may be ascribed to the different sensitivity to high-frequent alternative model. However, their powers are all acceptable. As to the alternative  $H_{14}$ , both of these two tests show the popular sigmoidal shape power curves and  $\tilde{S}_n^{OPG}$  slightly outperforms  $\tilde{S}_n^{DEE}$ .

Figure 1 about here

In summary, the proposed tests  $\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$  can both control type I error very well and make an excellent power performance. The OPG-based test  $\tilde{S}_n^{OPG}$  is more powerful.

*Study 2:* The data are generated from the following models:

$$\begin{aligned} H_{21} : Y &= \beta_1^\top X + 1.5a(\beta_2^\top X)^3 + \varepsilon, \\ H_{22} : Y &= \beta_1^\top X + 0.3a\left\{4(\beta_2^\top X)^3 + (\beta_2^\top X)^2\right\} + \varepsilon, \end{aligned}$$

$$H_{23} : Y = \beta_1^\top X + 4a \exp(-\beta_2^\top X) + \varepsilon,$$

where  $\beta_1 = (1, \dots, 1)^\top / \sqrt{p}$ ,  $\beta_2 = (\underbrace{1, \dots, 1}_{p/2}, 0, \dots, 0)^\top / \sqrt{p/2}$ ,  $p = 4, 2$  and  $n = 100, 200$ . When  $a \neq 0$ , we have  $q = 2$  and  $B = (\beta_1, \beta_2)$ . The covariates  $X$  and the error  $\varepsilon$  is independently generated from the multivariate and univariate standard normal distributions, respectively. For all of cases, 10% of the responses are randomly replaced by observations from a nonlinear model  $Y = 5.5 \cos(3\pi\beta_1^\top X) + \varepsilon$ . We intend to apply these alternative models to examine the effect of dimensionality on the proposed tests  $\tilde{S}_n^{OPG}$ ,  $\tilde{S}_n^{DEE}$  and  $T_n^{WQ}$  considered by Wang and Qu (2007).

The simulation results with the alternatives  $H_{21}, H_{22}$  are reported in Table 2 for  $a = 0, 0.2, \dots, 1$  at the significance level  $\alpha = 0.05$ . From this table, we can observe that for all of cases we conduct, the three tests can control empirical sizes very well. Also, it is reasonable that the simulated powers of all of tests become higher with increasing of the parameter  $a$  and the tests are more powerful with larger sample size. It can be seen clearly that the tests  $\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$  work very well in power performance.  $T_n^{OPG}$  and  $T_n^{DEE}$  are not significantly affected by the dimension of  $X$ . However,  $T_n^{WQ}$  severely suffers from the dimensionality problem. When the dimension of  $X$  gets larger,  $T_n^{WQ}$  completely fails to detect the alternatives. Figure 2 reports the simulated power curves under  $H_{23}$  for different values of  $a$  and sample sizes  $n = 100, 200$ . The similar conclusion can be made.

Table 2 and Figure 2 about here

*Study 3:* We generate the data from the following models:

$$H_{31} : Y = \beta^\top X + 2a(\beta^\top X)^3 + \varepsilon,$$

where  $\beta = (1, \dots, 1)^\top / \sqrt{p}$  and  $p = 8, 12$ . Here, the covariate  $X = (X_1, \dots, X_p)^\top$  come from the multivariate normal distribution  $N(0, I_p)$  where  $I_p$  is a  $p \times p$  identity matrix and  $\varepsilon$  is from univariate standard normal distribution  $N(0, 1)$ . In this study, we want to examine two issues: One is whether the test  $T_n^{GWZ}$  can maintain empirical size when there exist some outlier values in responses and the other is whether RDREAM can have an ac-

ceptable power performance when there are not outliers. The ratio of responses randomly replaced by observations from a nonlinear model  $Y = 6 \exp(-|\beta^\top X|) + \varepsilon$  is denoted as  $\rho$ .

Figure 3 presents the empirical size curves or “significance trace” of the three tests for different values of the ratio  $\rho$  and sample sizes  $n = 100, 200$ . In this case, the ratio  $\rho = 0, 0.02, \dots, 0.1$ . From this figure, we can see that when  $\rho = 0$ , all of the three tests can control empirical sizes very well which are all close to the pre-specified significance level 0.05. However, with the increasing of the ratio  $\rho$ ,  $\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$  outperform the test  $T_n^{GWZ}$ . The simulation results indicate that the new tests are not affected by the outlier values and they are more robust, whereas  $T_n^{GWZ}$  fails to work when outliers exist.

Figure 3 about here

We display the simulated powers curves for different values of  $a$  in Figure 4. Here, we consider  $\rho = 0$  and  $n = 100, 200$ . In other words, there are no outliers in the data. The parameter  $a$  is set to be  $0, 0.2, \dots, 1$ . Based on this figure, we can see that compared with  $T_n^{GWZ}$ , it is anticipated that  $T_n^{GWZ}$  has higher powers since their test employs more value-information whereas the robust tests only utilize the rank-information of responses. However, the powers of the new tests are still acceptable.

Figure 4 about here

## 6 Real data analysis

We now apply RDREAM to a real data set collected from a HIV clinical trial. The HIV positive patients in this study were randomly divided into four groups to receive antiretroviral regimen: (i) ZDV; (ii) didanosine (ddi); (iii) ZDV+ddi and (iv) ZDV+zalcitabine. A more detailed description of this real data set can be found in Hammer et al. (1996). Many researchers have made use of this data set to illustrate their dimension reduction estimation methods and further to compare the treatment effects of monotherapy (say (i)) and combined therapy (say (ii)-(iv)), including Ding and Wang (2011), Guo et al (2014) and Hu et al (2010). Recently, Niu et al (2015a) analyzed



this data set to test whether the nonparametric component is a partial linear regression function. The conclusion arrived was that a linear regression model would be proper for this data set.

In this dataset, there are 746 male patients who had not received antiretroviral therapy before the clinical trial and our study focuses on 473 patients who had observations in the variable CD4 cell counts at  $96 \pm 5$  weeks post therapy among them. Further, based on the way of therapy, we divide the dataset into two subsets: the first dataset has 105 male patients receiving monotherapy, say (i) and the second dataset contains 368 patients receiving combined therapies, say (ii)-(iv). For each dataset, the response variable  $Y$  is CD4 cell counts at  $96 \pm 5$  weeks post therapy and the four covariates are CD4 cell counts at baseline ( $X_1$ ), CD4 cell counts at  $20 \pm 5$  weeks ( $X_2$ ), CD8 cell counts at baseline ( $X_3$ ) and CD8 cell counts at  $20 \pm 5$  weeks ( $X_4$ ). For ease of explanation, all the covariates are standardized separately and the responses are centered.

It is our interest to test whether the data  $(Y, X)$  can be fitted with linear regression models where  $X = (X_1, X_2, X_3, X_4)^\top$ , that is,

$$\begin{aligned} H_{10} : E(Y|X) &= \beta_1^\top X \text{ for some } \beta_1 \in R^4, \\ H_{11} : E(Y|X) &= m(B_1^\top X) \neq \beta_1^\top X \text{ for any } \beta_1 \in R^4 \end{aligned} \quad (6.1)$$

for the first dataset and

$$\begin{aligned} H_{20} : E(Y|X) &= \beta_2^\top X \text{ for some } \beta_2 \in R^4, \\ H_{21} : E(Y|X) &= m(B_2^\top X) \neq \beta_2^\top X \text{ for any } \beta_2 \in R^4 \end{aligned} \quad (6.2)$$

for the second dataset, respectively. The same kernel function and bandwidth are adopted as simulation section. With our proposed RDREAM, we can obtain that for the first dataset, the  $p$ -values for tests  $\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$  are 0.696 and 0.678, respectively and the corresponding  $p$ -values for the second dataset are 0.645 and 0.361. All of these results indicate that the original two datasets can be fitted by linear regression models.

To investigate the influence of outlier values in the response space, we artificially replace the first 5% responses of each dataset by an outlying value 800. With these two new datasets, we apply our proposed methods to test (6.1) and (6.2), respectively. As to the first new dataset, the  $p$ -values for

$\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$  are 0.546 and 0.780, respectively. The corresponding  $p$ -values are 0.543 and 0.151 for the second dataset. The similar conclusion with original data can be made, which indicates that our test methods are robust.

We next add  $c_1 = (3, 4, 5, 6)$  copies of  $(800, 168, 174, 605, 640)$  to the first original dataset and  $c_2 = (12, 15, 18, 21)$  copies of  $(400, 370, 373, 739, 606)$  to the second original dataset, respectively, to carry out the same tests. Here, the maximum contamination rate is 5%. The  $p$ -values of these tests are listed in Table 3, which all suggest linear regression modelling.

Table 3 about here

## Appendix. Proofs of theorems

The following conditions are required for proving the theorems in Section 3.

- (C1) The joint probability density function  $f(z, y)$  of  $(Z, Y)$  is bounded. Both errors  $e_i$  and  $\varepsilon_i$  have bounded probability density functions.
- (C2) The density function  $f(B^\top X)$  of  $B^\top X$  on support  $\mathcal{Z}$  exists and has two bounded derivatives and satisfies

$$0 < \inf f(z) < \sup f(z) < 1.$$

- (C3) The kernel function  $\mathcal{K}(\cdot)$  is a bounded, derivative and symmetric probability density function and all the moments of  $\mathcal{K}(\cdot)$  exist. The bandwidth satisfies  $nh^2 \rightarrow \infty$ ,  $nh^{\hat{q}} \rightarrow \infty$ ,  $\int \mathcal{K}(u)du = 1$ .
- (C4) There exists an estimator  $\hat{\alpha}$  such that under the null hypothesis,  $\sqrt{n}(\hat{\alpha} - \alpha) = O_p(1)$ , where  $\alpha = (\beta, \theta)$  and under the local alternative sequences,  $\sqrt{n}(\hat{\alpha} - \tilde{\alpha}) = O_p(1)$ , where  $\alpha$  and  $\tilde{\alpha}$  are both interior points of  $\Theta$ , a compact and convex set.
- (C5) Denote  $\alpha = (\beta, \theta)^\top$  and there exists a positive continuous function  $G(x)$  such that  $\forall \alpha_1, \alpha_2$ ,  $|g(x, \alpha_1) - g(x, \alpha_2)| \leq G(x)|\alpha_1 - \alpha_2|$ .
- (C6) The matrix  $E\{\nabla m(B^\top X)\nabla m(B^\top X)^\top\}$  is positive definite where  $\nabla m(\cdot) = m'(\cdot)$  denotes the gradient of the function  $m(\cdot)$ .

(C7)  $\mathcal{M}_n(s)$  has the following expansion:

$$\mathcal{M}_n(s) = \mathcal{M}(s) + E_n\{\psi(X, Y, s)\} + R_n(s),$$

where  $E_n(\cdot)$  denotes the average over all sample points,  $E\{\psi(X, Y, s)\} = 0$  and  $E\{\psi^2(X, Y, s)\} < \infty$ .

(C8)  $\sup_s \|R_n(s)\|_F = o_p(n^{-1/2})$ , where  $\|\cdot\|_F$  denotes the Frobenius norm of a matrix.

**Remark 5.** *The conditions (C1) and (C5) are necessary for the robustness of Hampel influence function. Conditions (C2) – (C4) are needed for ensuring the asymptotic normality of our test statistic and the consistency of the parameter estimators, where condition (C3) is the common requisite for the kernel density estimation problem. Condition (C6) is assumed for OPG and (C7) – (C8) are for DEE.*

The following lemmas are used to prove the theorems in Section 3. We first give the proof of Lemma 1 in Section 2.

*Proof of Lemma 1.* Under the null hypothesis  $H_0$  and fixed alternative hypothesis  $H_n$ , the consistency for OPG-based estimate  $\hat{q} \rightarrow q$  as  $n \rightarrow \infty$  has been proved in Lemma 1 of Niu et al. (2015b). In the following, we only give the proof of DEE-based estimate and the same conditions in Theorem 4 of Zhu et al. (2010) are adopted.

Based on Theorem 2 in Zhu et al. (2010), under some conditions designed by them, it can be shown that  $\mathcal{M}_{n,n} - \mathcal{M} = O_p(n^{-1/2})$ . Further, the root-n consistency of the eigenvalues of  $\mathcal{M}_{n,n}$  is retained, that is,  $\hat{\lambda}_i - \lambda_i = O_p(n^{-1/2})$ . Note that when  $l \leq q$ ,  $\lambda_l > 0$  and for  $l > q$ , we have  $\lambda_l = 0$ .  $c = 1/\sqrt{nh}$  in our paper is recommended, thus, when  $nh \rightarrow \infty, h \rightarrow 0$ , we have  $1/\sqrt{n} = o(c)$  and  $c = o(1)$ . For  $1 \leq l < q$ ,

$$\begin{aligned} \frac{\hat{\lambda}_{q+1} + c}{\hat{\lambda}_q + c} - \frac{\hat{\lambda}_{l+1} + c}{\hat{\lambda}_l + c} &= \frac{\lambda_{q+1} + c + O_p(\frac{1}{\sqrt{n}})}{\lambda_q + c + O_p(\frac{1}{\sqrt{n}})} - \frac{\lambda_{l+1} + c + O_p(\frac{1}{\sqrt{n}})}{\lambda_l + c + O_p(\frac{1}{\sqrt{n}})} \\ &= \frac{c + O_p(\frac{1}{\sqrt{n}})}{\lambda_q + c + O_p(\frac{1}{\sqrt{n}})} - \frac{\lambda_{l+1} + c + O_p(\frac{1}{\sqrt{n}})}{\lambda_l + c + O_p(\frac{1}{\sqrt{n}})} \Rightarrow -\frac{\lambda_{l+1}}{\lambda_l} < 0. \end{aligned}$$

When  $l > q$ ,

$$\begin{aligned} \frac{\hat{\lambda}_{q+1} + c}{\hat{\lambda}_q + c} - \frac{\hat{\lambda}_{l+1} + c}{\hat{\lambda}_l + c} &= \frac{\lambda_{q+1} + c + O_p(\frac{1}{\sqrt{n}})}{\lambda_q + c + O_p(\frac{1}{\sqrt{n}})} - \frac{\lambda_{l+1} + c + O_p(\frac{1}{\sqrt{n}})}{\lambda_l + c + O_p(\frac{1}{\sqrt{n}})} \\ &= \frac{c + O_p(\frac{1}{\sqrt{n}})}{\lambda_q + c + O_p(\frac{1}{\sqrt{n}})} - \frac{c + O_p(\frac{1}{\sqrt{n}})}{c + O_p(\frac{1}{\sqrt{n}})} \Rightarrow -1 < 0. \end{aligned}$$

Therefore, we can conclude that under the null hypothesis  $H_0$  and under the fixed alternative hypothesis (1.2),  $\hat{q} \rightarrow q$  as  $n \rightarrow \infty$ , which completes the proof.  $\square$

The proof of Lemma 2 in Section 3 is given as follows.

*Proof of Lemma 2.* Under the local alternative hypothesis (3.2) with  $C_n = n^{-1/2}h^{-1/4}$ , the proof of OPG-based  $\hat{q} \rightarrow 1$  as  $n \rightarrow \infty$  has been given in Lemma 1 of Niu et al (2015b). We only state the proof for DEE-based estimate.

From the proof of Theorem 2 in Guo et al (2015), it is shown that under the local alternative hypothesis,  $\mathcal{M}_{n,n} - \mathcal{M} = O_p(C_n)$ . Further, we can get  $\hat{\lambda}_i - \lambda_i = O_p(C_n)$ .

Thus, Note that  $\lambda_1 > 0$  and for any  $l > 1$ , we have  $\lambda_l = 0$ . Consequently, under the condition that  $C_n = o(c)$  and  $c = o(1)$ ,

$$\begin{aligned} \frac{\hat{\lambda}_2 + c}{\hat{\lambda}_1 + c} - \frac{\hat{\lambda}_{l+1} + c}{\hat{\lambda}_l + c} &= \frac{\lambda_2 + c + O_p(C_n)}{\lambda_1 + c + O_p(C_n)} - \frac{\lambda_{l+1} + c + O_p(C_n)}{\lambda_l + c + O_p(C_n)} \\ &= \frac{c + O_p(C_n)}{\lambda_1 + c + O_p(C_n)} - \frac{c + O_p(C_n)}{c + O_p(C_n)} \Rightarrow -1 < 0. \end{aligned}$$

Thus under the local alternative (3.2), Lemma 1 holds. The proof of Lemma 2 is finished.  $\square$

**Lemma 3.** *Given the conditions (C1)-(C8) in the Appendix, we have*

$$nh^{1/2}(V_n - V_n^*) \xrightarrow{p} 0,$$

where  $\xrightarrow{p}$  represents convergence in probability and

$$V_n^* = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K}_h\{\hat{B}(\hat{q})^\top(x_i - x_j)\} [H(e_i) - \frac{1}{2}] [H(e_j) - \frac{1}{2}], \quad (\text{A.1})$$

here,  $\mathcal{K}_h(\cdot) = \mathcal{K}(\cdot/h)/h^{\hat{q}}$ .

*Proof of Lemma 3.* We first decompose  $V_n - V_n^*$  as

$$\begin{aligned}
V_n - V_n^* &= \frac{1}{n^3(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l=1}^n \sum_{k=1}^n \mathcal{K}_h\{\hat{B}(\hat{q})^\top(x_i - x_j)\} \\
&\quad \times [I(\hat{e}_l \leq \hat{e}_i) - I(e_l \leq e_i)][I(\hat{e}_k \leq \hat{e}_j) - I(e_k \leq e_j)] \\
&\quad + \frac{2}{n^3(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l=1}^n \sum_{k=1}^n \mathcal{K}_h\{\hat{B}(\hat{q})^\top(x_i - x_j)\} \\
&\quad \times [I(\hat{e}_l \leq \hat{e}_i) - I(e_l \leq e_i)][I(e_k \leq e_j) - \frac{n+1}{2n}] \\
&\quad + \frac{1}{n^3(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l=1}^n \sum_{k=1}^n \mathcal{K}_h\{\hat{B}(\hat{q})^\top(x_i - x_j)\} \\
&\quad \times \left\{ [I(e_l \leq e_i) - \frac{n+1}{2n}][I(e_k \leq e_j) - \frac{n+1}{2n}] \right. \\
&\quad \left. - [H(e_i) - \frac{1}{2}][H(e_j) - \frac{1}{2}] \right\} \\
&=: A_1 + A_2 + A_3. \tag{A.2}
\end{aligned}$$

Since  $e_i = y_i - g(\beta^\top x_i, \theta)$ , we further have  $\hat{e}_i - e_i = g(\beta^\top x_i, \theta) - g(\hat{\beta}^\top x_i, \hat{\theta})$ . Let  $\alpha = (\beta, \theta)^\top$  and  $L(\hat{\alpha})$  as

$$L(\hat{\alpha}) = \max_{1 \leq i \leq n} \sum_{l=1, l \neq i}^n |I(\hat{e}_l \leq \hat{e}_i) - I(e_l \leq e_i)|.$$

Denote  $\Omega = \{\alpha^* : \sqrt{n}|\alpha^* - \alpha_0| \leq \delta\}$  for  $\delta = O(1)$  and  $t(x_i, x_l, \alpha, \alpha^*) = [g(\beta^\top x_i, \theta) - g(\beta^{*\top} x_i, \theta^*)] - [g(\beta^\top x_l, \theta) - g(\beta^{*\top} x_l, \theta^*)]$ , then

$$\begin{aligned}
\sup_{\alpha^* \in \Omega} |L(\alpha^*)| &= \sup_{\alpha^* \in \Omega} \sum_{l=1, l \neq i}^n |I(e_l - e_i \leq t(x_i, x_l, \alpha, \alpha^*)) - I(e_l \leq e_i)| \\
&\leq \sup_{\alpha^* \in \Omega} \sum_{l=1, l \neq i}^n I(|e_l - e_i| \leq |t(x_i, x_l, \alpha, \alpha^*)|) \\
&\leq \sum_{l=1, l \neq i}^n I(|e_l - e_i| \leq Cn^{-1/2}), \tag{A.3}
\end{aligned}$$

where  $C$  is a generic positive constant. From the above derivation, we can obtain that, conditional on  $e_i$ ,  $I(|e_l - e_i| \leq Cn^{-1/2}), l \neq i$  are iid Bernoulli

random variables with  $O(n^{-1/2})$  order success probability. Using Bernstein's inequality leads to  $P\{\sum_{l=1, l \neq i}^n I(|e_l - e_i| \leq Cn^{-1/2}) \geq Cn^{1/2} | e_i\} \leq \exp(-Cn^{1/2})$ . Unconditionally, we still have that

$$P\left\{\sum_{l=1, l \neq i}^n I(|e_l - e_i| \leq Cn^{-1/2}) \geq Cn^{1/2}\right\} \leq \exp(-Cn^{1/2}).$$

Further,

$$P\left\{\max_{1 \leq i \leq n} \sum_{l=1, l \neq i}^n I(|e_l - e_i| \leq Cn^{-1/2}) \geq Cn^{1/2}\right\} \leq n \exp(-Cn^{1/2}).$$

When  $n \rightarrow \infty$ , for a positive constant  $C$ ,  $n \exp(-Cn^{1/2}) \rightarrow 0$ . Therefore,  $P\{\max_{1 \leq i \leq n} \sum_{l=1, l \neq i}^n I(|e_l - e_i| \leq Cn^{-1/2}) < Cn^{1/2}\} = 1$ . Then,

$$\max_{1 \leq i \leq n} \sum_{l=1, l \neq i}^n I(|e_l - e_i| \leq Cn^{-1/2}) < Cn^{1/2} = O_p(n^{1/2}). \quad (\text{A.4})$$

Combining (A.3) and (A.4), we can gain the following useful probability bound

$$L(\hat{\alpha}) = \max_{1 \leq i \leq n} \sum_{l=1, l \neq i}^n |I(\hat{e}_l \leq \hat{e}_i) - I(e_l \leq e_i)| = O_p(n^{1/2}).$$

An application of the formula (A.4) and for the term  $A_1$  in (A.2), we have

$$\begin{aligned} nh^{1/2}|A_1| &\leq \frac{h^{1/2}}{n^2(n-1)h^{\hat{q}}} \max_{1 \leq i \leq n} \sum_{l=1, l \neq i}^n |I(\hat{e}_l \leq \hat{e}_i) - I(e_l \leq e_i)| \\ &\quad \times \max_{1 \leq j \leq n} \sum_{k=1, k \neq j}^n |I(\hat{e}_k \leq \hat{e}_j) - I(e_k \leq e_j)| \\ &\quad \times \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K}\{\hat{B}(\hat{q})^\top (x_i - x_j)/h\} \\ &= \frac{Ch^{1/2}}{n(n-1)h^{\hat{q}}} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K}\{\hat{B}(\hat{q})^\top (x_i - x_j)/h\} \\ &= \frac{Ch^{1/2}}{n(n-1)h^{\hat{q}}} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K}\{B^\top (x_i - x_j)/h\} \end{aligned}$$

$$\begin{aligned}
& + \frac{Ch^{1/2}}{n(n-1)h^{\hat{q}}} \sum_{i=1}^n \sum_{j \neq i}^n \left[ \mathcal{K}\{\hat{B}(\hat{q})^\top(x_i - x_j)/h\} - \mathcal{K}\{B^\top(x_i - x_j)/h\} \right] \\
& =: C_1(A_{11} + A_{12}), \tag{A.5}
\end{aligned}$$

where  $C_1$  is a positive constant. Let  $Z = B^\top X$ . As to the term  $A_{11}$ , the term

$$\frac{1}{n(n-1)h^{\hat{q}}} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K}\{B^\top(x_i - x_j)/h\}$$

is an U-statistic with the kernel as  $H_n(z_1, z_2) = h^{-\hat{q}}\mathcal{K}\{(z_1 - z_2)/h\}$ . In order to apply the theory for non-degenerate U-statistic (Serfling 1980),  $E[H_n(z_1, z_2)^2] = o(n)$  is needed. It can be verified that

$$\begin{aligned}
E[H_n(z_1, z_2)^2] &= E\{E[H_n(z_1, z_2)^2|z_1, z_2]\} \\
&= \int \frac{1}{h^{2\hat{q}}} \mathcal{K}^2\left(\frac{z_1 - z_2}{h}\right) p(z_1) p(z_2) dz_1 dz_2 \\
&= \int \frac{1}{h^{2\hat{q}}} \mathcal{K}^2(u) p(z_1) p(z_1 - hu) (-h^{\hat{q}}) dz_1 du \\
&= -\frac{1}{h^{\hat{q}}} \int \mathcal{K}^2(u) p^2(z_1) dz_1 du + o(1) \\
&= O\left(\frac{1}{h^{\hat{q}}}\right), \tag{A.6}
\end{aligned}$$

where  $p(\cdot)$  is denoted as the probability density function. With the condition  $nh^{\hat{q}} \rightarrow \infty$ , we have  $E[H_n(z_1, z_2)^2] = O(1/h^{\hat{q}}) = o(n)$ . The condition of lemma 3.1 of Zheng (1996) is satisfied and we have  $A_{11} = h^{1/2}E[H_n(z_1, z_2)] + o_p(1)$ , where  $E[H_n(z_1, z_2)] = O(1)$ . Therefore we can obtain that  $A_{11} = O_p(h^{1/2}) = o_p(1)$ . Denote

$$A_{12}^* = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n h^{1/2-\hat{q}} \mathcal{K}'\{\tilde{B}^\top(x_i - x_j)/h\} (x_i - x_j)^\top \times \frac{\hat{B}(\hat{q}) - B}{h},$$

where  $\tilde{B}$  lies between  $B$  and  $\hat{B}$ . Then for the term  $A_{12}$  in (A.5), we have

$$A_{12} = A_{12}^* + o_p(A_{12}^*).$$

Similar to  $A_{11}$ , the following term

$$\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n h^{-\hat{q}} \mathcal{K}'\{\tilde{B}^\top(x_i - x_j)/h\} (x_i - x_j)^\top$$

can be regarded as an U-statistic. It can be similarly shown that the term is the order of  $O_p(h)$ . As  $\|\hat{B}(\hat{q}) - B\|_2 = O_p(1/\sqrt{n})$  and under the condition  $n \rightarrow \infty, h \rightarrow 0$ , we can obtain that  $A_{12} = o_p(1)$ . Towards to (A.5), we have

$$nh^{1/2}|A_1| \leq o_p(1).$$

Similarly, we can derive that  $nh^{1/2}|A_i| = o_p(1)$  for  $i = 2, 3$ . Combining with the formula (A.2), it can be concluded that

$$nh^{1/2}(V_n - V_n^\star) \xrightarrow{p} 0,$$

which completes the proof of Lemma 3.  $\square$

In the following, we give the proof for Theorem 1.

*Proof of Theorem 1.* From Lemma 3, we know that the limiting distributions for  $nh^{1/2}V_n$  and  $nh^{1/2}V_n^\star$  are the same. Thus, we just need to derive the asymptotic property of  $nh^{1/2}V_n^\star$ . The term  $V_n^\star$  in (A.1) can be decomposed as

$$\begin{aligned} V_n^\star &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K}_h\{B^\top(x_i - x_j)\} [H(e_i) - \frac{1}{2}] [H(e_j) - \frac{1}{2}] \\ &\quad + \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n [H(e_i) - \frac{1}{2}] [H(e_j) - \frac{1}{2}] [\mathcal{K}_h\{\hat{B}(\hat{q})^\top(x_i - x_j)\} - \mathcal{K}_h\{B^\top(x_i - x_j)\}] \\ &=: V_{n1}^\star + V_{n2}^\star, \end{aligned}$$

where  $\mathcal{K}_h(\cdot) = \mathcal{K}(\cdot/h)/h^{\hat{q}}$ .

For the term  $V_{n1}^\star$ , it is a U-statistic, since we always assume that the dimension of  $B^\top X$  is fixed in our paper. Under the null hypothesis,  $H(e_i)$ ,  $i = 1, \dots, n$  follows a uniform distribution on  $(0, 1)$ ,  $q = 1$  and  $\hat{q} \rightarrow 1$ . An application of Theorem 1 in Zheng (1996), it is not difficult to derive the asymptotic normality:  $nh^{1/2}V_{n1}^\star \Rightarrow N(0, Var)$ , where

$$Var = \frac{1}{72} \int \mathcal{K}^2(u) du \int p^2(z) dz$$

with  $Z = B^\top X$ .

Denote

$$\tilde{V}_{n2}^\star = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K}' \left\{ \frac{\tilde{B}^\top(x_i - x_j)}{h} \right\} (x_i - x_j)^\top [H(e_i) - \frac{1}{2}] [H(e_j) - \frac{1}{2}] \cdot \frac{\hat{B}(\hat{q}) - B}{h},$$



where  $\tilde{B}$  lies between  $\hat{B}$  and  $B$ . An application of Taylor expansion yields

$$V_{n2}^* = \tilde{V}_{n2}^* + o_p(\tilde{V}_{n2}^*).$$

Because the kernel  $\mathcal{K}(\cdot)$  is spherical symmetric, the following term can be considered as an U-statistic:

$$\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\tilde{q}}} \mathcal{K}'\{\tilde{B}^\top(x_i - x_j)/h\} (x_i - x_j)^\top [H(e_i) - \frac{1}{2}] [H(e_j) - \frac{1}{2}].$$

Further note that

$$\begin{aligned} & E\left\{ \frac{1}{h^{\tilde{q}}} \mathcal{K}'\{\tilde{B}^\top(x_i - x_j)/h\} (x_i - x_j)^\top [H(e_i) - \frac{1}{2}] [H(e_j) - \frac{1}{2}] |x_i, y_i \right\} \\ &= E\left[ E\left\{ \frac{1}{h^{\tilde{q}}} \mathcal{K}'\{\tilde{B}^\top(x_i - x_j)/h\} (x_i - x_j)^\top [H(e_i) - \frac{1}{2}] [H(e_j) - \frac{1}{2}] |x_i, y_i, x_j \right\} |x_i, y_i \right] \\ &= E\left\{ \frac{1}{h^{\tilde{q}}} \mathcal{K}'\{\tilde{B}^\top(x_i - x_j)/h\} (x_i - x_j)^\top [H(e_i) - \frac{1}{2}] \cdot E[H(e_j) - \frac{1}{2} |x_j] |x_i, y_i \right\} = 0. \end{aligned}$$

Thus the above U-statistic is degenerate. Similar as the derivation of  $V_{n1}^*$ , together with  $\|\hat{B}(\hat{q}) - B\|_2 = O_p(1/\sqrt{n})$  and  $1/nh^2 \rightarrow 0$ , we have  $nh^{1/2}V_{n2}^* = o_p(1)$ . Therefore, under the null hypothesis  $H_0$ , we can conclude that  $nh^{1/2}V_n^* \Rightarrow N(0, Var)$ . Based on Lemma 3, we have  $nh^{1/2}V_n \Rightarrow N(0, Var)$ .

An estimate of  $Var$  can be defined as

$$\widehat{Var} = \frac{1}{72n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\tilde{q}}} \mathcal{K}^2\left\{ \frac{\hat{B}(\hat{q})^\top(x_i - x_j)}{h} \right\}.$$

Since the proof is rather straightforward, we then only give a brief description. Using a similar argument as that for Lemma 3, we can get

$$\widehat{Var} = \frac{1}{72n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\tilde{q}}} \mathcal{K}^2\left\{ \frac{B^\top(x_i - x_j)}{h} \right\} + o_p(1).$$

The consistency can be derived through U-statistic theory. The proof for Theorem 1 is finished.  $\square$

The proof for Theorem 2 is given as follows.

*Proof of Theorem 2.* Under the global alternative  $H_n$  in (1.2), we have  $e_i = m(B^\top x_i) + \varepsilon_i - g(\tilde{\beta}^\top x_i, \tilde{\theta})$ . Together with Lemma 3, it can be obtained

that  $V_n = V_n^* + o_p(1)$ , where  $V_n^*$  can be rewritten as

$$\begin{aligned}
V_n^* &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{\hat{B}(\hat{q})^\top (x_i - x_j)}{h} \right\} \left[ H(m(B^\top x_i) + \varepsilon_i - g(\tilde{\beta}^\top x_i, \tilde{\theta})) - \frac{1}{2} \right] \\
&\quad \left[ H(m(B^\top x_j) + \varepsilon_j - g(\tilde{\beta}^\top x_j, \tilde{\theta})) - \frac{1}{2} \right] \\
&= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{B^\top (x_i - x_j)}{h} \right\} \left[ H(m(B^\top x_i) + \varepsilon_i - g(\tilde{\beta}^\top x_i, \tilde{\theta})) - \frac{1}{2} \right] \\
&\quad \left[ H(m(B^\top x_j) + \varepsilon_j - g(\tilde{\beta}^\top x_j, \tilde{\theta})) - \frac{1}{2} \right] + \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \left[ H(m(B^\top x_i) + \varepsilon_i \right. \\
&\quad \left. - g(\tilde{\beta}^\top x_i, \tilde{\theta})) - \frac{1}{2} \right] \left[ H(m(B^\top x_j) + \varepsilon_j - g(\tilde{\beta}^\top x_j, \tilde{\theta})) - \frac{1}{2} \right] \left[ \mathcal{K}_h \{ \hat{B}(\hat{q})^\top (x_i - x_j) \} \right. \\
&\quad \left. - \mathcal{K}_h \{ B^\top (x_i - x_j) \} \right] \\
&= V_{n3}^* + V_{n4}^*,
\end{aligned}$$

For the term  $V_{n3}^*$ , it is a standard U-statistic with

$$H_n(x_i, x_j) = \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{B^\top (x_i - x_j)}{h} \right\} l(x_i) l(x_j),$$

where  $l(x) = [H\{m(B^\top x) + \varepsilon. - g(\tilde{\beta}^\top x, \tilde{\theta})\} - 1/2]$ . Similar to the proof of (A.6), when  $nh^{\hat{q}} \rightarrow \infty$ , we can derive that  $E[H^2(x_i, x_j)] = o(n)$  and the condition of Lemma 3.1 in Zheng (1996) can be shown to be satisfied. We further cacluate

$$\begin{aligned}
E[H_n(x_i, x_j)] &= E\{E[H_n(x_i, x_j)|x_i, x_j]\} \\
&= \frac{1}{h^{\hat{q}}} \int \mathcal{K} \left\{ \frac{z_i - z_j}{h} \right\} \tilde{l}(z_i) \tilde{l}(z_j) p(z_i) p(z_j) dz_i dz_j \\
&= \frac{1}{h^{\hat{q}}} \int \mathcal{K}(u) \tilde{l}(z_j + hu) \tilde{l}(z_j) p(z_j + hu) p(z_j) \times h^{\hat{q}} du dz_j \\
&= \int \tilde{l}^2(z_j) p^2(z_j) dz_j + o(1) \\
&= E[l^2(X)^2 p(X)] + o(1).
\end{aligned}$$

where  $Z = B^\top X$ . Therefore,  $V_{n3}^* = E[l^2(X)^2 p(X)] + o_p(1) =: C_2$ , here,  $C_2$  is a positive constant.

As to the term  $V_{n4}^*$ , similarly as the term  $V_{n2}^*$ , we have

$$V_{n4}^* = \tilde{V}_{n4}^* + o_p(\tilde{V}_{n4}^*),$$

where,

$$\tilde{V}_{n4}^* = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K}' \left\{ \frac{\tilde{B}^\top(x_i - x_j)}{h} \right\} (x_i - x_j)^\top l(x_i) l(x_j) \cdot \frac{\hat{B}(\hat{q}) - B}{h},$$

here,  $\tilde{B}$  lies between  $B$  and  $\hat{B}$ . Similarly as the derivation of  $V_{n3}^*$ , together with  $\|\hat{B}(\hat{q}) - B\|_2 = O_p(1/\sqrt{n})$ , when  $nh^{\hat{q}} \rightarrow \infty$ , we have  $V_{n4}^* = O_p(h) \cdot O_p(1/\sqrt{n}) \cdot (1/h) = o_p(1)$ .

Based on the above analysis, we can derive that  $V_n = C_2 + o_p(1)$  and  $nh^{1/2}V_n \Rightarrow \infty$  in probability, which completes the proof of the global alternative situation.

We now consider the situation of local alternative  $H_{1n}$  in (3.2). Based on Lemma 3, we have  $V_n = V_n^* + o_p(1)$ . In this situation,  $e_i = C_n m(B^\top x_i) + \varepsilon_i$ . Therefore,  $V_n^*$  can be decomposed as

$$\begin{aligned} V_n^* &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{\hat{B}(\hat{q})^\top(x_i - x_j)}{h} \right\} \left[ H(C_n m(B^\top x_i) + \varepsilon_i) - \frac{1}{2} \right] \\ &\quad \left[ H(C_n m(B^\top x_j) + \varepsilon_j) - \frac{1}{2} \right] \\ &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{B^\top(x_i - x_j)}{h} \right\} \left[ H(C_n m(B^\top x_i) + \varepsilon_i) - \frac{1}{2} \right] \\ &\quad \left[ H(C_n m(B^\top x_j) + \varepsilon_j) - \frac{1}{2} \right] + \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \left[ H(C_n m(B^\top x_i) + \varepsilon_i) \right. \\ &\quad \left. - \frac{1}{2} \right] \left[ H(C_n m(B^\top x_j) + \varepsilon_j) - \frac{1}{2} \right] \left[ \mathcal{K}_h \{ \hat{B}(\hat{q})^\top(x_i - x_j) \} - \mathcal{K}_h \{ B^\top(x_i - x_j) \} \right] \\ &= V_{n5}^* + V_{n6}^*, \end{aligned} \tag{A.7}$$

For the term  $V_{n5}^*$ , taking a Taylor expansion of  $H(C_n m(B^\top x_i) + \varepsilon_i)$  around  $C_n = 0$ , we have

$$H(C_n m(B^\top x_i) + \varepsilon_i) = H(\varepsilon_i) + C_n h(\varepsilon_i) m(B^\top x_i) + o_p(C_n^2).$$

Then, the term  $V_{n5}^*$  can be decomposed as

$$\begin{aligned}
V_{n5}^* &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{B^\top(x_i - x_j)}{h} \right\} \left[ H(\varepsilon_i) - \frac{1}{2} \right] \left[ H(\varepsilon_j) - \frac{1}{2} \right] \\
&\quad + 2C_n \left\{ \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{B^\top(x_i - x_j)}{h} \right\} \left[ H(\varepsilon_i) - \frac{1}{2} \right] h(\varepsilon_j) m(B^\top x_j) \right\} \\
&\quad + C_n^2 \left\{ \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{B^\top(x_i - x_j)}{h} \right\} h(\varepsilon_i) m(B^\top x_i) h(\varepsilon_j) m(B^\top x_j) \right\} + o_p(C_n^2) \\
&= D_1 + 2C_n D_2 + C_n^2 D_3 + o_p(C_n^2).
\end{aligned}$$

Under the local alternative hypothesis,  $\hat{q} \rightarrow 1$  can be obtained. For the term  $D_1$ , similarly to the proof of the term  $V_{n1}^*$  in Theorem 1, we can show that  $nh^{1/2}D_1 \Rightarrow N(0, Var)$ , where

$$Var = \frac{1}{72} \int \mathcal{K}^2(u) du \int p^2(z) dz$$

with  $Z = B^\top X$ . As to the term  $D_2$ , similarly as the proof of Lemma 3.3b in Zheng (1996), it can be obtained that  $D_2 = O_p(1/\sqrt{n})$ . When  $C_n = n^{-1/2}h^{-1/4}$ , we have  $nh^{1/2}C_n D_2 = O_p(h^{1/2})$ . Turn to the term  $D_3$ , similarly to the proof of  $V_{n3}^*$  in our Theorem 2, we have  $D_3 = E[h^2(\varepsilon)m^2(B^\top X)p(X)] + o_p(1)$ . Further,  $nh^{1/2}C_n^2 D_3 = E[h^2(\varepsilon)m^2(B^\top X)p(X)] + o_p(1)$ . Therefore,

$$nh^{1/2}V_{n5}^* \Rightarrow N(\mu, Var),$$

where  $\mu = E[h^2(\varepsilon)m^2(B^\top X)p(X)]$ .

As to the term  $V_{n6}^*$ , just similarly as the proof of the term  $V_{n2}^*$  in our Theorem 1, it can be gotten that  $nh^{1/2}V_{n6}^* = o_p(1)$ .

Combining Lemma 3 and the formula (A.7), under the local alternative, we have  $nh^{1/2}V_n \Rightarrow N(\mu, Var)$ .

The proof of Theorem 2 is finished.  $\square$

The verification for the formula (4.2) is as follows.

*Verification of (4.2).* Let  $f_t$  and  $u$  be the probability density functions of  $F_t$  and  $U$ . Recall  $F_t = F + tU$ , then  $dF_t = dF + tdU$  and  $f_t = f + tu$ , we further have

$$L(t) = \int \int \left[ H(y - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \times \left( \int \left[ H(y_1 - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \right.$$

$$\begin{aligned}
& \times f_t(z, y_1) dy_1) dF_t(z, y), \\
= & \int \int \left[ H(y - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \times \left( \int \left[ H(y_1 - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \right. \\
& \times f(z, y_1) dy_1) dF(z, y) + t \int \int \left[ H(y - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \\
& \times \left( \int \left[ H(y_1 - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \times u(z, y_1) dy_1 \right) dF(z, y) \\
& + t \int \int \left[ H(y - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \times \left( \int \left[ H(y_1 - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \right. \\
& \times f(z, y_1) dy_1) dU(z, y) + t^2 \int \int \left[ H(y - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \\
& \times \left( \int \left[ H(y_1 - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \times u(z, y_1) dy_1 \right) dU(z, y) \\
=: & L_1(t) + L_2(t) + L_3(t) + L_4(t).
\end{aligned}$$

For the first term  $L_1(t)$ , we have

$$\begin{aligned}
\frac{dL_1(t)}{dt} = & \int \int \frac{d}{dt} H(y - g(\beta^\top x(F_t), \theta(F_t))) \times \left( \int \left[ H(y_1 - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \right. \\
& \times f(z, y_1) dy_1) dF(z, y) + \int \int \left[ H(y - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \\
& \times \left( \int \left( \frac{d}{dt} H(y_1 - g(\beta^\top x(F_t), \theta(F_t))) \times f(z, y_1) dy_1 \right) dF(z, y).
\end{aligned}$$

Since  $\int [H(y - g(\beta^\top x(F_t), \theta(F_t))) - 1/2] dF(y|z) = 0$ , we have that

$$\frac{dL_1(t)}{dt} \Big|_{t=0} = 0.$$

Further,

$$\begin{aligned}
\frac{d^2 L_1(t)}{dt^2} \Big|_{t=0} = & 2 \int \int \dot{H}(y - g(\beta^\top x(F), \theta(F))) \times \left( \int \left( \dot{H}(y_1 - g(\beta^\top x(F), \theta(F))) \right. \right. \\
& \times f(z, y_1) dy_1) dF(z, y).
\end{aligned}$$

Similarly, it is not difficult to obtain that  $\frac{dL_i(t)}{dt} \Big|_{t=0} = 0$ ,  $i = 2, 3, 4$  and  $\frac{1}{2} \frac{d^2 L_2(t)}{dt^2} \Big|_{t=0}$ ,  $i = 2, 3, 4$  are equal to other three terms in the formula (4.2), which completes the proof.  $\square$

## References

- [1] Bierens, H.J. (1990) A consistent conditional moment test of functional form. *Econometrica*, **58**, 1443-1458.
- [2] Bierens, H.J. and Ploberger W. (1997). Asymptotic theory of integrated conditional moment test. *Econometrica*, **65**, 1129-1151.
- [3] Cook, R.D. and Weisberg, S. (1991). Discussion of “Sliced inverse regression for dimension reduction” by K.C. Li. *Journal of the American Statistical Association*, **86**, 328-332.
- [4] Ding, X.B. and Wang, Q.H. (2011). Fusion-refinement procedure for dimension reduction with missing response at random. *Journal of the American Statistical Association*, **106**, 1193-1207.
- [5] Escanciano, J.C. (2006a). Goodness-of-fit tests for linear and nonlinear time series models. *Journal of the American Statistical Association*, **101**, 531-541.
- [6] Escanciano, J.C. (2006b). A consistent diagnostic test for regression models using projections. *Econometric Theory*, **22**, 1030-1051.
- [7] Escanciano, J.C. (2007). Model checks using residual marked empirical processes. *Statistica Sinica*, **17**, 115-138.
- [8] Fan, J.Q., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and the wilks phenomenon. *The Annals of Statistics*, **29**, 153-193.
- [9] Feng, L., Zou, C.L., Wang, Z.J. and Zhu, L.X. (2015). Robust comparison of regression curves. *Test*, **24**, 185-204.
- [10] Fernholz, L.T. (1983). *von Mises calculus for statistical functionals*. New York: Springer-Verlag.
- [11] González-Manteiga, W. and Crujeiras, R.M. (2013). An updated review of Goodness-of-Fit tests for regression models. *Test*, **22**, 361-411.

- [12] Guo, X., Wang, T., Xu, W.L. and Zhu, L.X. (2014). Dimension reduction with missing response at random. *Computational Statistics and Data Analysis*, **69**, 228-242.
- [13] Guo, X., Wang, T. and Zhu, L.X. (2015). Model checking for parametric single index models: A dimension-reduction model-adaptive approach. *Journal of the Royal Statistical Society: Series B*. In Press.
- [14] Hammer, S.M., et al. (1996). A trial comparing nucleotide monotherapy with combined therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, **335**, 1081-1090.
- [15] Hampel, F.R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, **69**, 383-393.
- [16] Härdle, W. (1992). *Applied nonparametric regression*. Cambridge, U.K.: Cambridge University Press.
- [17] Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, **21**, 1296-1947.
- [18] Heritier, S. and Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association*, **89**, 897-904.
- [19] Horowitz, J.L. and Spokoiny V.G. (2001). An adaptive, rate-optimal test of a parametric meanregression model against a nonparametric alternative. *Econometrica*, **69**, 599-631.
- [20] Hu, Z.H., Follmann, D.A. and Qin, J. (2010). Semiparametric dimension reduction estimation for mean response with missing data. *Biometrika*, **97**, 305-319.
- [21] Huber, P.J. (1985). Projection pursuit. *The Annals of Statistics*, **13**, 435-475.
- [22] Koul, H.L. and Ni, P.P. (2004). Minimum distance regression model checking. *Journal of Statistical Planning and Inference*, **119**, 109-141.

- [23] Lavergne, P. and Patilea, V. (2008). Breaking the curse of dimensionality in nonparametric testing. *Journal of Econometrics*, **143**, 103-122.
- [24] Lavergne, P. and Patilea, V. (2012). One for all and all for one: regression checks with many regressors. *Journal of Business & Economic Statistics*, **30**, 41-52.
- [25] Li, K.C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, **86**, 316-327.
- [26] Li, Y.C. and Zhu, L.X. (2007). Asymptotics for sliced average variance estimation. *The Annals of Statistics*, **35**, 41-69.
- [27] Markatou, M. and Hettmansperger. (1990). Robust bounded-influence tests in linear models. *Journal of American Statistical Association*, **85**, 187-190.
- [28] Markatou, M. and Manos, G. (1996). Robust tests in nonlinear regression models. *Journal of Statistical Planning and Inference*, **55**, 205-217.
- [29] Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability and Its Applications*, **10**, 186-196.
- [30] Niu, C.Z., Guo, X., Xu, W.L. and Zhu, L.X. (2015a). Checking nonparametric component for partial linear regression model with missing response. *Journal of Statistical Planning and Inference*. In press.
- [31] Niu, C.Z., Guo, X. and Zhu, L.X. (2015b). Enhancements of nonparametric generalized likelihood ratio test: Bias-correction and dimension reduction. *arXiv: 1507.06090*.
- [32] Schrader, R.M. and Hettmansperger, T.P. (1980). Robust analysis of variance based upon a likelihood ratio criterion. *Biometrika*, **67**, 93-101.
- [33] Sen, P.K. (1982). On  $M$ -tests in linear models. *Biometrika*, **69**, 245-248.
- [34] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley, New York.



- [35] Stone, C.J. (1980). Optimal rates of convergence for nonparametric estimators. *Annals of Statistics*, **8**, 1348-1360.
- [36] Stute, W. (1997). Nonparametric model checks for regression. *Annals of Statistics*, **25**, 613-641.
- [37] Stute, W., González-Manteiga, W. and Presedo-Quindimil, M. (1998a). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, **93**, 141-149.
- [38] Stute, W., Xu, W. L. and Zhu, L. X. (2008). Model Diagnosis for Parametric Regression in High Dimensional Spaces. *Biometrika*. 95. 1-17.
- [39] Stute, W., Thies, S. and Zhu, L.X. (1998b). Model checks for regression: An innovation process approach. *Annals of Statistics*. **26**, 1916-1934.
- [40] Stute, W. and Zhu, L.X. (2002). Model checks for generalized linear models. *Schandinavian Journal of Statistics*, **29**, 535-545.
- [41] Van Keilegom, I., González-Manteiga, W. and Sánchez Sellero, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *Test*, **17**, 401-415.
- [42] Wang, L. and Qu, A.N. (2007). Robust tests in regression models with omnibus alternatives and bounded influence. *Journal of the American Statistical Association*, **102**, 347-358.
- [43] Watson, G.S. (1964). Smoothing regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, **26**, 359-372.
- [44] Whang, Y.J. (2000). Consistent bootstrap tests of parametric regression functions. *Journal of Econometrics*, **98**, 27-46.
- [45] Xia, Q, Xu, W.L. and Zhu, L.X. (2015). Consistently determining the number of factors in multivariate volatility modelling. *Statistica Sinica*, **25**, 1025-1044.
- [46] Xia, Y.C., Tong, H., Li, W.K. and Zhu, L.X. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society: Series B*, **64**, 363-410.

- [47] Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, **75**, 263-289.
- [48] Zhu, L.X. (2003). Model checking of dimension-reduction type for regression. *Statistica Sinica*, **13**, 283-296.
- [49] Zhu, L. X. and An, H. Z.(1992). A test for nonlinearity in regression models. *Journal of Mathematics*, **4**, 391-397. (Chinese).
- [50] Zhu, L. X. and Li, R. Z. (1998). Dimension-reduction type test for linearity of a stochastic model. *Acta Mathematicae Applicatae Sinica*. **14**, 165-175.
- [51] Zhu, L.P., Wang, T., Zhu, L.X. and Ferré, L. (2010). Sufficient dimension reduction through discretization-expectation estimation. *Biometrika*, **97**, 295-304.

Table 1: Empirical sizes and powers of  $\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$  for  $H_{10}$  v.s.  $H_{11}$  and  $H_{12}$  at the significance level  $\alpha = 0.05$  with  $p = 8$ .

$\varepsilon$		$a$	$n = 60$		$n = 100$		$n = 200$	
			$\tilde{S}_n^{OPG}$	$\tilde{S}_n^{DEE}$	$\tilde{S}_n^{OPG}$	$\tilde{S}_n^{DEE}$	$\tilde{S}_n^{OPG}$	$\tilde{S}_n^{DEE}$
$H_{11}$	$\varepsilon \sim \ln N(0, 0.25^2)$	0	0.048	0.058	0.049	0.044	0.050	0.046
		0.2	0.095	0.127	0.129	0.133	0.208	0.163
		0.4	0.161	0.169	0.284	0.177	0.629	0.417
		0.6	0.306	0.170	0.562	0.342	0.925	0.856
		0.8	0.379	0.255	0.725	0.553	0.976	0.972
		1.0	0.508	0.317	0.856	0.717	0.989	0.998
	$\varepsilon \sim N(0, 1)$	0	0.048	0.057	0.051	0.049	0.052	0.047
		0.2	0.081	0.107	0.118	0.144	0.196	0.149
		0.4	0.164	0.138	0.267	0.197	0.614	0.428
		0.6	0.250	0.174	0.537	0.347	0.894	0.837
		0.8	0.420	0.244	0.725	0.550	0.983	0.960
		1.0	0.479	0.340	0.850	0.702	0.991	0.994
$H_{12}$	$\varepsilon \sim \ln N(0, 0.25^2)$	0	0.045	0.061	0.045	0.052	0.052	0.046
		0.2	0.064	0.103	0.092	0.142	0.318	0.372
		0.4	0.156	0.171	0.351	0.407	0.835	0.852
		0.6	0.296	0.297	0.606	0.624	0.976	0.979
		0.8	0.378	0.406	0.764	0.753	0.993	0.995
		1.0	0.498	0.498	0.860	0.868	0.998	0.999
	$\varepsilon \sim N(0, 1)$	0	0.051	0.059	0.056	0.055	0.050	0.046
		0.2	0.068	0.076	0.104	0.142	0.277	0.340
		0.4	0.136	0.169	0.350	0.351	0.797	0.829
		0.6	0.273	0.251	0.586	0.583	0.978	0.945
		0.8	0.377	0.382	0.759	0.739	0.997	0.994
		1.0	0.482	0.473	0.870	0.859	0.999	1.000

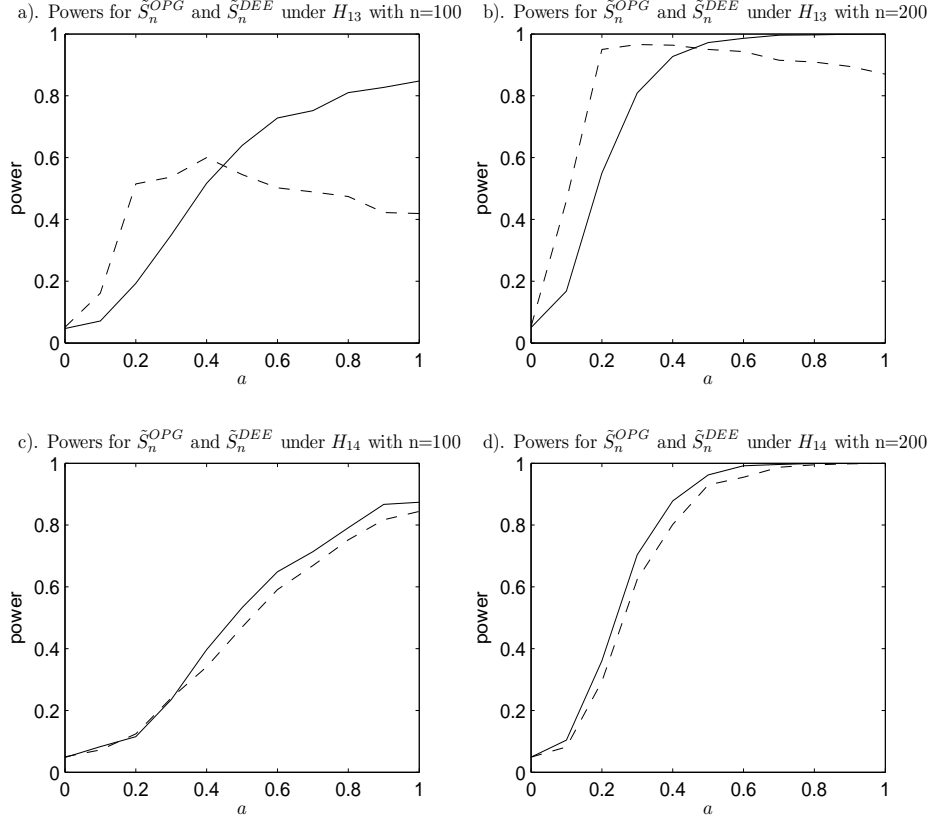


Figure 1: Empirical sizes and powers of  $\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$  for  $H_{10}$  v.s.  $H_{13}, H_{14}$  at the significance level  $\alpha = 0.05$  with  $p = 8$ ,  $X \sim N(0, I_p)$  and  $\varepsilon \sim \ln N(0, 0.25^2)$ . In four plots, the solid line and the dash line are for  $\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$ , respectively.

Table 2: Empirical sizes and powers of  $\tilde{S}_n^{OPG}$ ,  $\tilde{S}_n^{DEE}$  and  $T_n^{WQ}$  for  $H_{20}$  v.s.  $H_{21}$  and  $H_{22}$  at the significance level  $\alpha = 0.05$ .

	$p$	$a$	$n = 100$			$n = 200$		
			$\tilde{S}_n^{OPG}$	$\tilde{S}_n^{DEE}$	$T_n^{WQ}$	$\tilde{S}_n^{OPG}$	$\tilde{S}_n^{DEE}$	$T_n^{WQ}$
$H_{21}$	$p = 4$	0	0.045	0.055	0.048	0.051	0.057	0.045
		0.2	0.145	0.112	0.055	0.389	0.164	0.056
		0.4	0.479	0.273	0.058	0.871	0.648	0.061
		0.6	0.708	0.501	0.062	0.976	0.874	0.067
		0.8	0.833	0.649	0.067	0.990	0.956	0.071
		1.0	0.893	0.733	0.070	0.997	0.982	0.075
	$p = 2$	0	0.047	0.057	0.046	0.053	0.056	0.051
		0.2	0.165	0.115	0.071	0.405	0.226	0.129
		0.4	0.536	0.352	0.121	0.925	0.681	0.345
		0.6	0.801	0.527	0.192	0.986	0.909	0.559
		0.8	0.882	0.706	0.300	0.995	0.967	0.698
		1.0	0.940	0.802	0.344	0.992	0.989	0.798
$H_{22}$	$p = 4$	0	0.045	0.049	0.053	0.047	0.053	0.052
		0.2	0.105	0.073	0.055	0.243	0.144	0.056
		0.4	0.404	0.204	0.058	0.746	0.496	0.060
		0.6	0.592	0.379	0.060	0.932	0.759	0.065
		0.8	0.771	0.554	0.063	0.981	0.912	0.070
		1.0	0.842	0.636	0.067	0.996	0.956	0.073
	$p = 2$	0	0.051	0.052	0.059	0.041	0.055	0.049
		0.2	0.123	0.097	0.064	0.275	0.160	0.101
		0.4	0.429	0.230	0.108	0.801	0.531	0.252
		0.6	0.684	0.405	0.170	0.970	0.800	0.440
		0.8	0.819	0.597	0.214	0.994	0.927	0.605
		1.0	0.897	0.706	0.287	0.996	0.964	0.717

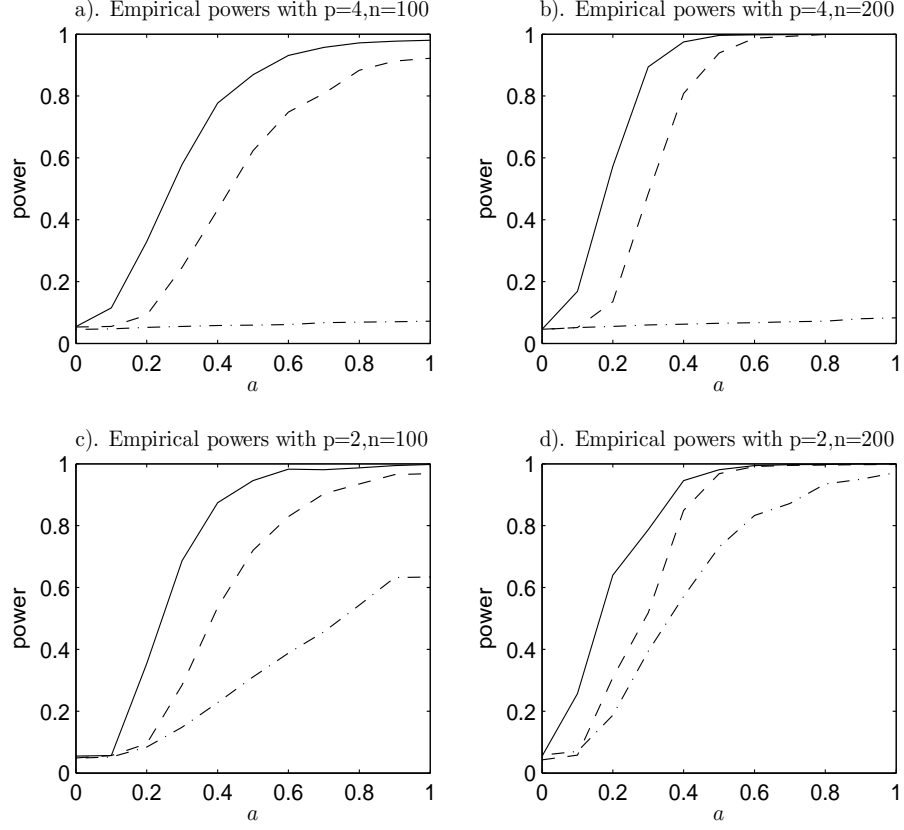


Figure 2: Empirical sizes and powers of  $\tilde{S}_n^{OPG}$ ,  $\tilde{S}_n^{DEE}$  and  $T_n^{WQ}$  for  $H_{20}$  v.s.  $H_{23}$  at the significance level  $\alpha = 0.05$  with  $X \sim N(0, I_p)$  and  $\varepsilon \sim N(0, 1)$ . In four plots, the solid line, the dash line and the dash-dotted line are for  $\tilde{S}_n^{OPG}$ ,  $\tilde{S}_n^{DEE}$  and  $T_n^{WQ}$ , respectively.

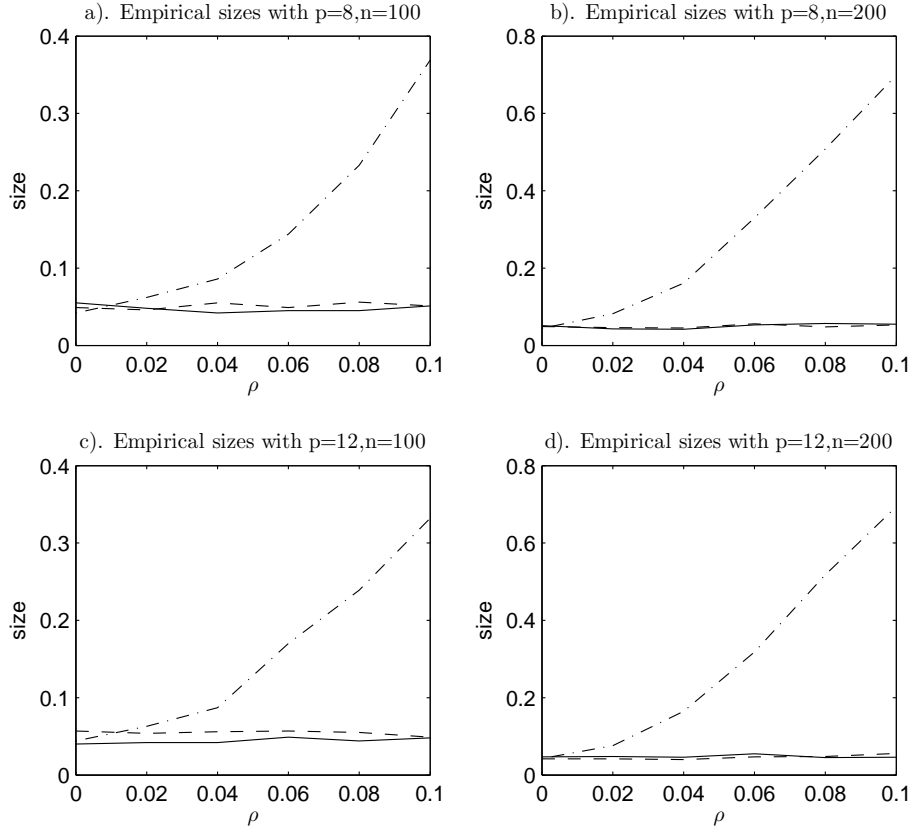


Figure 3: Empirical sizes of  $\tilde{S}_n^{OPG}$ ,  $\tilde{S}_n^{DEE}$  and  $T_n^{GWZ}$  for  $H_{30}$  at the significance level  $\alpha = 0.05$  with  $X \sim N(0, I_p)$ ,  $\varepsilon \sim N(0, 1)$  and different values of  $\rho$ . In four plots, the solid line, the dash line and the dash-dotted line are for  $\tilde{S}_n^{OPG}$ ,  $\tilde{S}_n^{DEE}$  and  $T_n^{GWZ}$ , respectively.

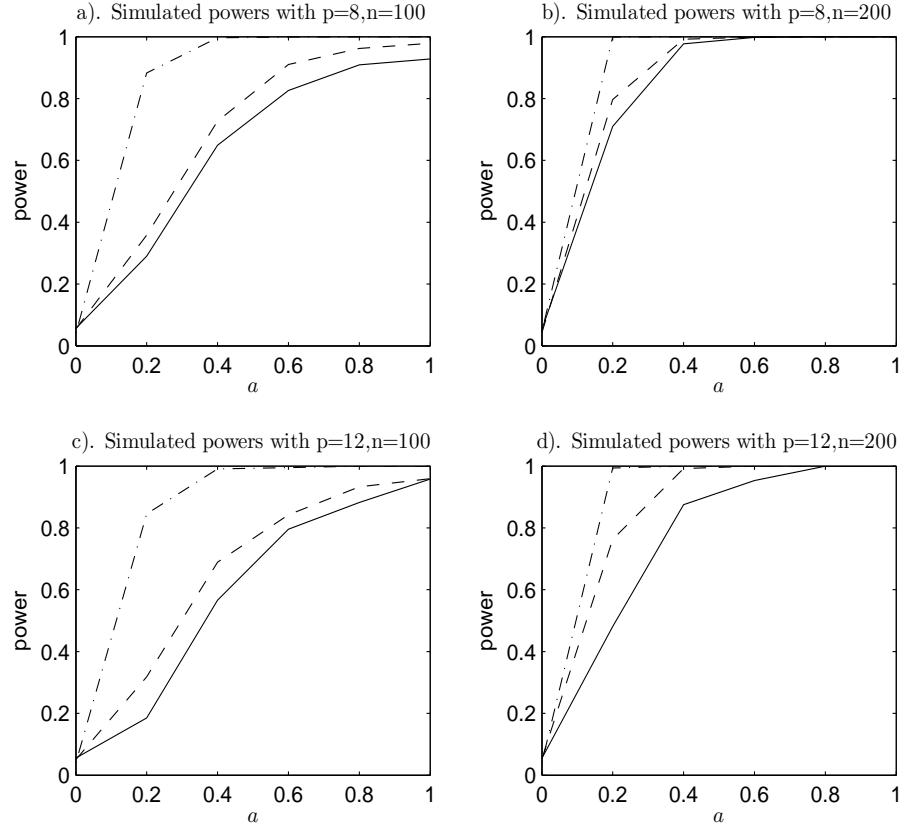


Figure 4: Simulated powers of  $\tilde{S}_n^{OPG}$ ,  $\tilde{S}_n^{DEE}$  and  $T_n^{GWZ}$  for  $H_{30}$  v.s.  $H_{31}$  at the significance level  $\alpha = 0.05$  with  $X \sim N(0, I_p)$  and  $\varepsilon \sim N(0, 1)$ . In four plots, the solid line, the dash line and the dash-dotted line are for  $\tilde{S}_n^{OPG}$ ,  $\tilde{S}_n^{DEE}$  and  $T_n^{GWZ}$ , respectively.



Table 3:  $p$ -values of  $\tilde{S}_n^{OPG}$  and  $\tilde{S}_n^{DEE}$  for the real data analysis.

$c_1$	The first subset		$c_2$	The second subset	
	$\tilde{S}_n^{OPG}$	$\tilde{S}_n^{DEE}$		$\tilde{S}_n^{OPG}$	$\tilde{S}_n^{DEE}$
3	0.273	0.934	12	0.802	0.528
4	0.400	0.527	15	0.587	0.755
5	0.686	0.235	18	0.374	0.909
6	0.919	0.119	21	0.207	0.532